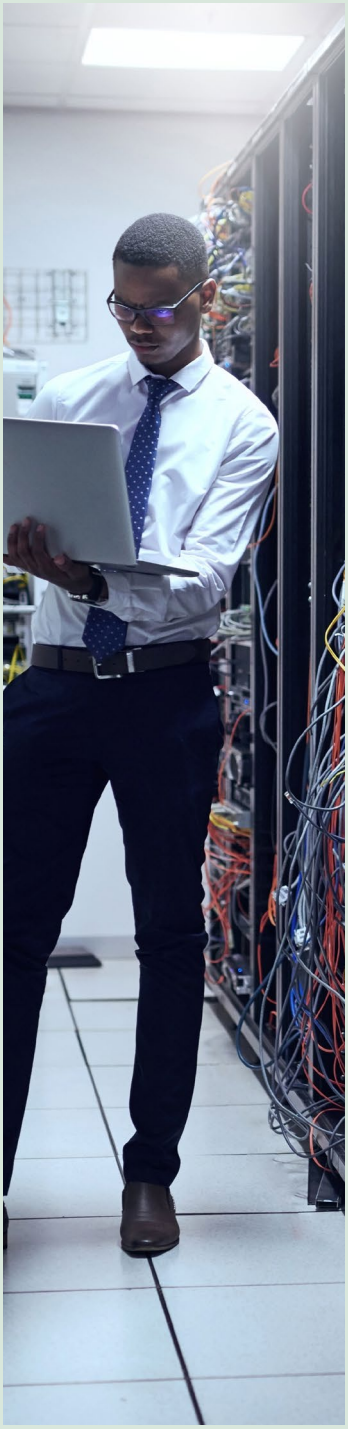
Infosys topaz

WHITE PAPER



CLOUD COST OPTIMIZATION

Abstract

Cloud computing has seen a paradigm shift in software industry over last few years. As per Google Cloud Brand Pulse Survey conducted in 2023, 41.4% of global business and tech leaders aim to increase their investment in cloud-based services. It is safe to say that cloud will be a key driver of organizational growth across verticals. The cloud computing has revolutionized multiple industries like retail, finance, social media, education, health insurance, telecom etc. Furthermore, it is set to change the landscape of new and emerging industries like AI, edge computing and more.

Although, migrating to cloud is a very remunerative solution, it poses challenges of its own if not done optimally. As of 2023 Statista survey, conducted globally, around 82 percent of respondents found the biggest challenge of using cloud computing were related to managing cloud spend. Organizations waste an average of around 32 % of their total cloud cost. The main reason deemed for the same is lack of proper assessment, planning and outlining a continuous process. Optimal cloud architecture requires implementing custom solutions based on 'in-depth' performance analysis, cost saving strategies and best practices.

This paper outlines Infosys' holistic approach towards cloud cost management. It helps enterprises identify and address inefficiencies resulting in significant cloud cost reduction.

Infosys®
Navigate your next

Table of Contents

Factors contributing to cloud cost

With the continuous acceleration of cloud usage, cloud spend has become a significant portion of IT budget. Ensuring efficiency by putting in right tools, processes and people has become very critical in realizing business values. Infosys recommends a proactive approach to understanding and optimizing cloud spend.

Three critical elements of cloud spending are as follows:

1. Compute: Computational power for big data requires advanced and niche computing technologies. AI/ML workload have raised this requirement significantly.
2. Storage: Data storage – though usually inexpensive – can add to overall cloud costs when the data volumes are large.
3. Data Lakes: Ingestion and consumption of data on data lake platforms require the transfer of enormous amounts of data across the cloud network.

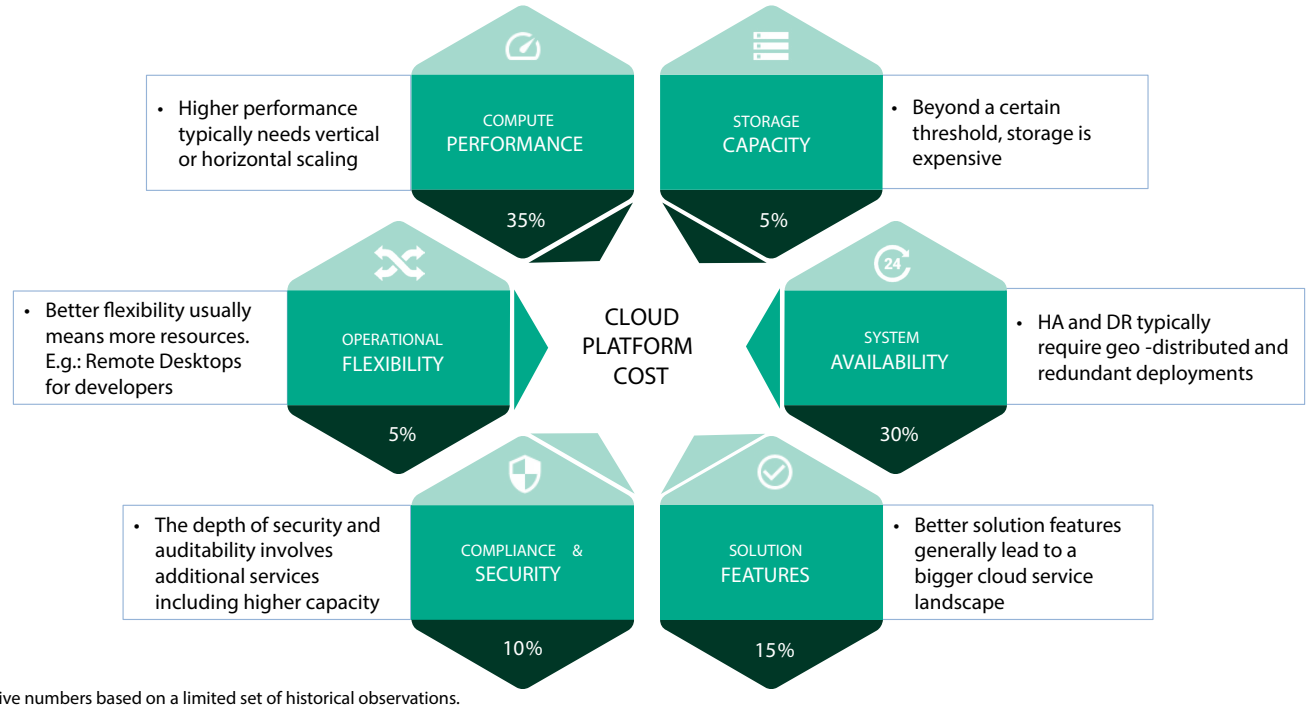
While off-the-shelf infrastructure pieces are available, it requires deep understanding of both the problem domain and myriad cloud infrastructure offering to re-architect traditional applications to be cloud ready.

Infosys has deep expertise in cloud computing technologies across multiple verticals. This help Infosys to provide industry specific cloud solutions as per customer's requirements. Moreover, the customer benefits from Infosys's abstracted best practices for both the cloud and the vertical.

Cost attributions : Observation over last few years

While observing the cloud spend by enterprise over the past 2 decades, we've concluded that about 35% of the cost is attributed to compute performance. This aligns with the general understanding that higher processing power requires both vertical and horizontal scaling. AI/ML processing is a new paradigm which requires not only high number of CPU cores but GPU processing as well. Around 30% of expenditure is assigned to enhancing system availability (SA). Constituents of SA such as high availability (HA) and disaster recovery (DR), require geo-distributed and redundant deployments. About 15% of cost is attributed to solution features – a larger cloud service landscape results from better solution features. The remaining 20% is attributable to compliance and security, storage capacity, and operational flexibility.

Cloud Data Platforms - Cost Attribution Per Historical Observations



* Indicative numbers based on a limited set of historical observations.

Factors contributing to cloud cost

With the continuous acceleration of cloud usage, cloud spend has become a significant portion of IT budget. Ensuring efficiency by putting in right tools, processes and people has become very critical in realizing business values. Infosys recommends a proactive approach to understanding and optimizing cloud spend.

Three critical elements of cloud spending are as follows:

- **Awareness:** Lack of awareness of OPEX (Operational expenditure) centric cost model native to the cloud. Lack of awareness/knowledge of the various cost components of the cloud services availed. Correspondingly, lack of visibility on the various configurations or plans within the service that can cater to multiple service-level agreements (SLAs), specifically for data platforms.
- **Lack of Assessment:** When it comes to deliberating cloud strategies and optimizing cloud costs, the phrase 'fit-for-purpose' is better suited than 'best-of-breed' as it requires an in-depth performance analysis based on some preassigned pre-decided metrics. Falling in for discounts and paying upfront/locking period without correct use-case can lead to higher cost while we intend to reduce the same with the discounts.
- **Lack of Planning:** A clear plan for cloud spends on data and non-data-centric aspects of the service, and the allotment of resources towards each of them are undefined. Optimization planning and aligning to best practices need to be planned before implementation.
- **Misalignment:** Lack of alignment between the solution's architecture and right/fundamental cloud functions, such as storage and decoupled computing.
- **Time to Market:** Finally, the timely implementation of cost governance measures is often absent. Once launched, cost optimization needs to be continuous and automated.

Few Best Practices for Optimizing Cloud Cost

Cost optimization best practices can be bucketed into following categories

Right-sizing

Selecting a virtual machine instance that enables a firm to meet all its basic workload requirements efficiently can result in monthly savings of up to 70%. A crucial aspect of capacity planning for storage-intensive workloads is the assessment of peak input/output operations per second (IOPS) storage requirements. By considering factors such as CPU, RAM, storage, and network utilization, right-sizeable instances can be identified. Based on insights from monitoring instances, they can be configured to scale up or down.

Elasticity

Make use of elasticity tools such as auto-scaling, load balancers etc, provided by vendor for varying data loads. Auto scaling provides users with an automated approach to increase/decrease compute, memory or networking resources based on traffic or utilization.

Good hygiene best practices:

Simple practices like pausing AWS services such as RDS, Redshift or Azure dedicated SQL pool(formerly known as SQL DW) compute resources when unengaged helps save money. Automations to shutdown resources at the end of the day also help save money.

Location

Re-evaluation replication level and redundancy across geographic locations also aids in reducing cloud costs. Replacing hot-hot implementations with hot-cold architecture depending on service-level agreements (SLAs) and disabling unrequired geo-redundancies are some ways of achieving cost-efficiency goals. Note. However, that replacing consistency is a fundamental architecture change and might need application changes as well.

Pricing models

Choosing cost-effective pricing options for different types of workloads can help organizations save a considerable amount of monetary allocation for cloud services. Choose "pay as you go model"(on demand instances) for non-steady, spiky, sporadic workloads, that charges according to actual usage or consumption. Use Reserved instances for known, steady-state workloads. Use spot instances for fault-tolerant, flexible and stateless workloads.

Class

Picking appropriate storage class, switching hot blob storage with cold blob storage for archived data, using edge nodes and cloud front distributions, introducing caching among others are some of the choices that can help improve the capacity of the cloud architecture.

Culture orientation

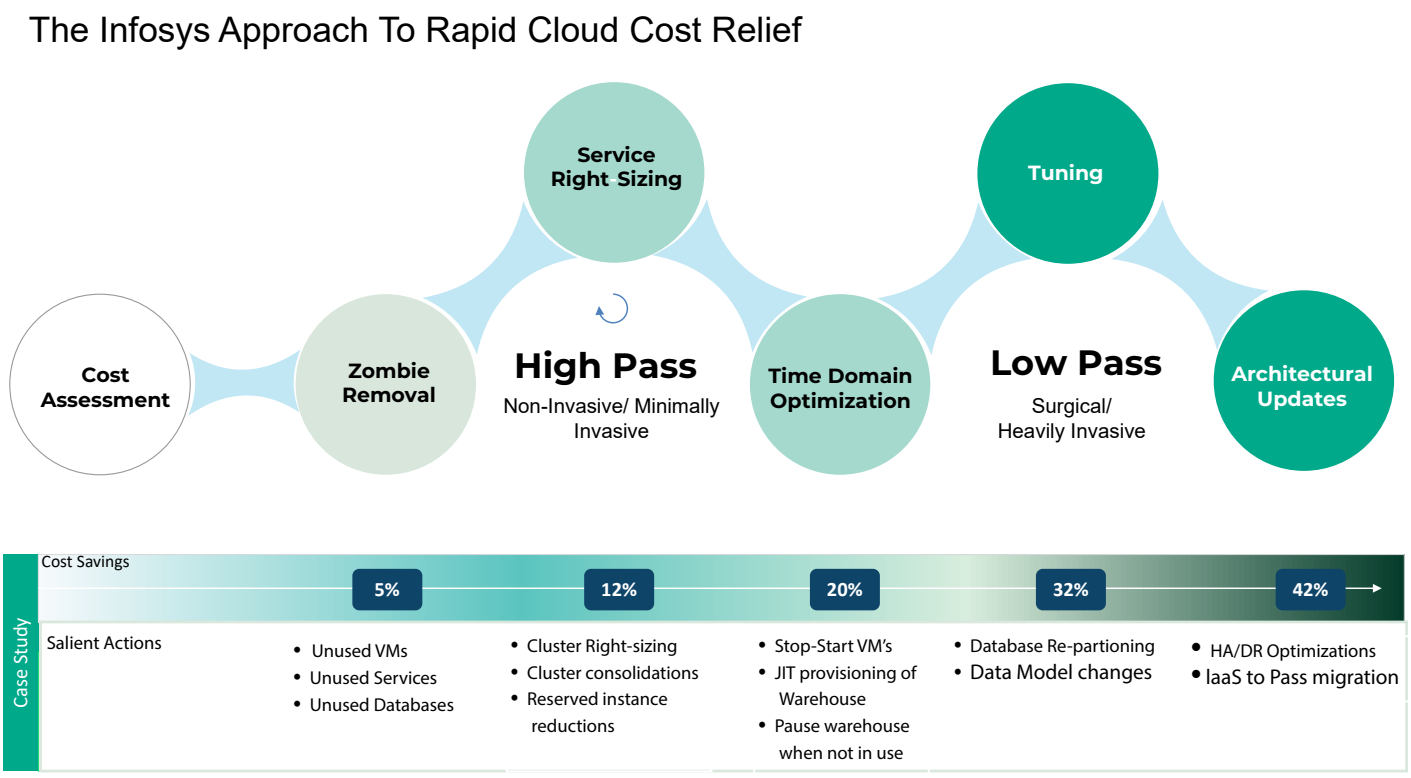
Establishing a culture of cost-consciousness across the organization contributes to the lowering of overall cloud costs and maximizes ROI on cloud investments. For example, analysis of spending on business units (BU), accounts, and services aids in identifying areas that require targeted cost optimization efforts. Examining instance utilization metrics, such as memory and network usage, helps discern underutilized instances. Identifying Peak load /Off peak loads timings, cost and usage. Adopting a serverless service model, decoupled architecture.

Consolidated billing provides institutions with a comprehensive overview of their total cloud spending.

Infosys' cloud cost management approach

Cloud cost optimization has a significant bearing on the cost benefits(ROI) achieved by the client. Infosys approach to the reduction of cloud costs is not only comprehensive but also effectuates a rapid decrease in expenditure. A thorough cost assessment followed by zombie removal, non-invasive, minimal-invasive and finally highly-invasive/surgical help reduce cost significantly.

Infosys' strategy has three levels: quick cost relief, invasive cost optimization, and cost avoidance.



Quick Cost relief :

It involves reducing costs by identifying low-hanging items using the Pareto or '80-20' principle. Identifying and terminating unused virtual machines, event hubs and databases can save up to 5 % of total cost. **Non Invasive/Minimal Invasive** practices like right sizing, aligning to best practices, applying trust advisor recommendations etc. can save up to 12% of total cost. Time Domain Optimizations involving stop-start vms, pausing databases etc. can save up to 20 % of cost.

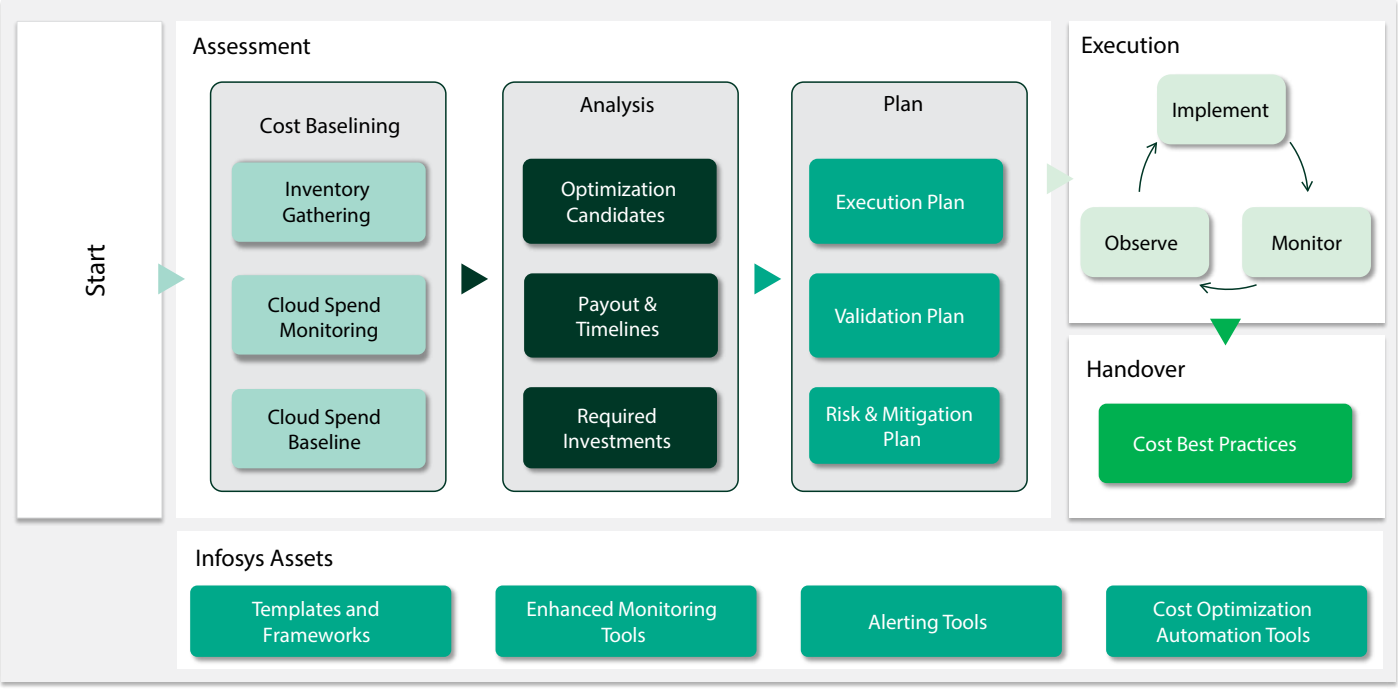
Invasive Cost optimization:

This process focuses on a detailed optimization process aimed at sustained cost reduction and addresses consumption, design, architecture, and licensing. If the configuration of the overall cloud framework requires large-scale revamping to achieve optimization, the architecture is upgraded/optimized. A cloud service's configuration, directly and indirectly, impacts its overall cost. A direct effect on cost can be observed in the case of provisioned capacity – the higher the provisioned capacity, the higher the costs. Indirect effects on cost arise from sub-optimal configuration, affecting overall cloud performance, including slower response times and higher latencies.

These changes are marked as surgical or heavily invasive measures. Re-Partitioning, Model changes etc. can save up to 32 % of total cost. Architectural Updates can save 42% of the total cost.

Invasive Cost Optimization is the end-to-end optimization of the client's cloud setup. It begins with **assessing** the existing processes and involves the various facets of cost baselining, analysis, and planning. This requires invasive intervention, particularly broad changes to application designs for cost optimization. An indicative percentage of cost reduced could be anywhere around 25 to 35 %.

End to End Cloud Cost Optimization Approach



Assessment is followed by the execution of the plan wherein the changes are implemented, Cost is monitored and observed. Finally customized architecture is handed over. Infosys' proprietary assets are vital in optimizing the client's cloud deployment overall. They include templates and frameworks, enhanced monitoring tools, alerting tools, and cost optimization automation tools.

Cost Avoidance

Infosys helps avoiding cost which entails the technical implementations and instrumentations for cost reduction in the long run via optimal cost best practices as mentioned above.

The type of procurement/ license agreement with the cloud vendor including EA/CSP, the discounts received from vendors, has an overall impact on the cloud cost. Infosys enables the beneficial alignment of procurement models with the organization's cost optimization goals.

This above comprehensive approach results in the delivery of personalized cloud solutions for organizations.

Been there, Done that

There are multiple success stories of clients who experienced beneficial cloud cost optimization with Infosys's engagement. For example, as a result of a 6-week initiative, a leading retail chain achieved annual savings of $1.53 million per year. The monthly cloud spending witnessed a shared reduction: from $235,000 to $95,000. Some of the core AWS services that were optimized included Elastic Compute Cloud (EC2), Elastic Block Store (EBS), and Elastic MapReduce (EMR). This was all achieved in 6 weeks.

Similarly, Infosys empowered a Fortune 500 automotive manufacturing company to optimize cloud consumption costs for their real-time analytics platform on Azure. A 47% reduction in cloud costs, amounting to $2.47 million in annual savings was achieved. During the 8-week-long project, cloud services such as CosmoDB, HDInsight, and Application Insights were also optimized.

Hence, Infosys – and its solutions – can be an ideal partner for organizations seeking to optimize their cloud costs.

Authored By:

Sarika Nandwani

For more information, contact askus@infosys.com

Infosys®
Navigate your next