LTIMindtree

Whitepaper

The Right Way to Approach Cloud Cost Optimization

Authored by:

Gaurav Goel

Senior Cloud Architect

Table of Contents

Abstract 3

Challenges 3

How to start? 3

What are your options for optimization? 4

How LTIMindtree can help you? 8

©LTIMindtree | Privileged and Confidential 2023

Abstract

As every IT geek, student, or professional knows, the cloud is the new elephant in the room, for a substantial portion of the IT expenditure is spent on hosting the IT infrastructure. Almost every organization, private or public, has either adopted the cloud already or is in the process of doing so.

Cloud adoption is a multi-phased exercise and the last of it is cost optimization and re-invention, which is a continuous and repeated cycle. As a cloud expert, you need to periodically assess your current cloud expenditure, find the areas where the spending is more than it should be, and re-invent your applications, if required, to leverage the true advantages of the cloud along with other technological changes to optimize the cost.

Cloud cost optimization exercise is a must because if not used the right way, you may end up spending over the allocated budget and miss one of the essential reasons for moving to the cloud.

Challenges

However, identifying the areas for cost optimization is not as easy or straightforward as it may seem. It requires you to have a very good understanding of the cloud environment along with the applications hosted, their architectures, resources in the cloud and their purpose, and many other pieces of information. Therefore, it is not a one-man job. It requires input from many cloud architects, application architects, security experts, etc.

Some of the cost optimization exercises can be easier and quicker to implement. Their results (cost effect) can also be seen in the AWS invoice almost immediately, while the others can be difficult. The time taken to implement them will be more and require much more planning and discussion with other stakeholders.

How to start?

Before starting the actual cost optimization exercise, you need to understand the current cost spending and identify some of the spending patterns and trends, such as:

1. Invoice per AWS service
2. Invoice per AWS account
3. Invoice per application
4. Service vs. application
5. Service vs. account
6. Application vs. account

©LTIMindtree | Privileged and Confidential 2023

What are your options for optimization?

Following are the ways you can review the current cloud spending and identify opportunities to optimize the cost:

Quickwins: The easiest and quickest way is to start with identifying the Quickwins, which are the resources that are no longer required or used seldom. Those resources can either be deleted or shut down/stopped/auto-parked. This will reduce the invoice immediately.

Also, as a best practice, any resource meant for proof of concept (PoC) or testing should be created in a separate sandbox or test account and deleted after serving the purpose. I recommend creating a governance policy to identify periodically and automatically the resources existing for more than a certain period, as applicable. This will ensure those resources are not overlooked and cleaned up when not required.

Instance rightsizing: Overprovisioning resources is a common mistake many organizations make. In fear of performance issues or lack of information/ knowledge available, many times organizations choose to go with the same resource configuration as on-premises while migrating resources to the cloud. This results in overprovisioning them because the on-prem infrastructure does not allow resources to scale up or down and they must be configured with a fixed size.

Analyze the usage pattern and resource utilization history for the last couple of months and identify the right configuration/size for the resource, including but not limited to Elastic compute cloud (EC2) and Relational Database Service (RDS) instance rightsizing, DynamoDB table read-write capacity, etc.

Auto-scaling: One of the true advantages of the cloud is the ability to autoscale the infrastructure up or down on demand. You can and you should design the applications to support autoscaling. Some third-party vendor applications do not support autoscaling, but you should always look for alternate ways, if possible, or at least design the in-house applications to support it. Autoscaling is something that can lower the total cost significantly and may not be exceedingly difficult to implement if the application supports it.

Therefore, whenever configuring the resource capacity for any resource in the cloud, see if there is an option to auto-scale it. If yes, choose it.

Instance auto-stop: The compute and database instances for non-production environments, especially the Dev and Test environments, may not require running 24x7 and, therefore, should be auto-stopped during non-business hours or should be started only on-demand. I have seen many organizations not having it in place and, therefore, wasting much money here. Once implemented,

©LTIMindtree | Privileged and Confidential 2023

the cost effects are huge and immediately visible. I recommend creating a governance policy to identify the instances periodically and automatically without auto-stop enabled/configured and enforce its implementation.

The solutions to auto-stop the instances based on schedule, tags, and other parameters I suppose, are readily available on most of the public cloud platforms such as AWS and Azure (I am sure of).

Object storage lifecycle management: While object storage such as AWS S3 and Azure Storage accounts provide a simpler way to manage object storage in the cloud along with redundancy, unlimited storage, and CDN capability features, I have seen organizations not paying attention to the unnecessary cost these services can add to their cloud spends. Therefore, it becomes particularly important to understand the cost implications of these services and implement the right policies and procedures to ensure the spending on these services is carefully observed.

You can and should implement the appropriate lifecycle rules to transfer the objects to lower-tier storage based on their use case and expiration based on the retention period. I recommend creating a governance policy to identify the object storage periodically and automatically without any lifecycle management rules and enforce its implementation.

Application architecture review: Now comes the challenging task and real game changer. As we all know, technology keeps evolving every decade, and the pace of change has only increased with the evolution and adoption of Microservices, Infrastructure as Code, Open-Source Software, Automation, and so on. Therefore, it becomes very important to review the architectures of our applications and identify the opportunities to reinvent them, which can help us leverage the best of new technologies and make sure there are improvements to the overall performance and availability of the application with minimum cost.\

Adopting modern technologies, especially DevOps and automation, can help us reduce the time to market and helps deliver products and updates to the customer easily and quickly giving you the edge in today's much more competitive world.

While reviewing the application architecture, focus on the following.

- Re-platform / Re-architect: See if you have chosen the right technologies for the use cases. A wrong choice may result in elevated license costs, support prices, maintenance costs, etc.
- Can open-source software be used instead?
- Opportunities to automate tasks/processes.
- All the earlier cost optimization options discussed above.

Consolidation: Resource consolidation is not something that may give a lot of cost benefits but can help reduce it in certain use cases. Also, it helps you greatly from an Operations perspective by reducing the number of resources to manage. Look for any consolidation opportunity while reviewing the resources in your cloud environment.

©LTIMindtree | Privileged and Confidential 2023

Go serverless: Serverless should be the way forward whenever designing new applications on the cloud or planning the rearchitecting. It rids you of the responsibility of managing IT infrastructure and focusing on your business. Often, it also helps reduce the total cost in the longer term. It takes you away from certain investments such as IT infrastructure management, technical resources required to manage them, skillset upgrades, etc. Such costs are not as visible as the others but impact the IT budget significantly.

Automation: The effect that automation can have on IT costs and efforts is unimaginable. Automation is the master key to unlocking the cost benefits of the cloud. If implemented the right way, it can save thousands of man-hours and dollars. It also reduces the chances of errors because it removes the human factor from the process. Oftentimes, it is not possible to calculate the direct savings but with detailed analysis, it can be done. As an architect (cloud or application), you should always look for opportunities to automate things.

In fact, automation is one of the ingredients required to leverage the true potential and advantages of cloud computing.

Spot instances: Spot instances are something that I have not seen many organizations use but can bring cloud spending significantly down. Spot instances are the spare hardware available with cloud providers which are available at cheaper rates, even up to 80-90% lower than the actual rates. They are a great fit for applications where abrupt interruption or termination is not an issue. If possible, the applications should be redesigned to leverage the spot instances for the huge cost savings they can provide.

The usage of spot instances should be encouraged and mandated, if possible, for the non-production environments. They can be a little challenging to implement and may require a bit of change to the application because of their possibility of abrupt termination but are worth every bit of effort because of the huge cost savings possible with them.

Reservation: While the on-premises infrastructure guarantees you the required capacity whenever required; it requires you to plan some upfront investments. On the other hand, the cloud takes you off the responsibility of upfront investment and the same level of capacity planning as on-premises. However, it does not guarantee the required capacity always because the cloud infrastructure is made available to other customers also and you may run into unavailability issues on rare occasions.

Reserved Instances allow you to reserve required Infrastructure with a cloud provider for a certain period. This way you are guaranteed the hardware availability whenever required. Reserved Instances can not only solve this problem, but they can also provide some significant cost savings from 20-70% as compared to the on-demand instances. Most public cloud providers make available the option to reserve required capacity for 1-year or 3-year terms and no-upfront, partial-upfront, and all-upfront payment options. The more you commit and the more you pay upfront, the more cost savings you get.

©LTIMindtree | Privileged and Confidential 2022

Planning for the reserved instance purchase requires a lot of planning and preparation because even if those commitments are not utilized, you still must pay for them. Therefore, before purchasing any reserved instances, do careful planning.

Savings plan: Savings plans are another saving option based on the commitment to the cloud provider. They are more flexible than the reserved instances and may cover a more significant number of resources under the savings comparatively because they apply to a lot more services rather than only EC2 and RDS instances (in AWS).

Just like the reserved instances, the savings plan purchase requires a lot of planning and preparation because even if those commitments are not utilized, you still have to pay for them.

While all the above exercises can be done whenever required, it is not enough to do it only once or seldom because you keep adding or changing the resources in the cloud and those changes always affect the cloud spending. It is recommended to create a process to do this exercise at least once a year and to mandate that all changes have gone through the Architectural review by the cloud experts. Cost is one of the pillars of the Well-Architected review in the cloud.

Cost optimization reviews should be a habit instead of a mandate.

©LTIMindtree | Privileged and Confidential 2023

How LTIMindtree can help you?

As I explained earlier, Automation is a real game changer and takes you off a load of manual efforts while also reducing the chances of human errors and saving money, of course. Doing the whole Cost Optimization review requires lots of manual effort and involvement. It is a long-running process; therefore, if you can find a way to automate it, nothing like it.

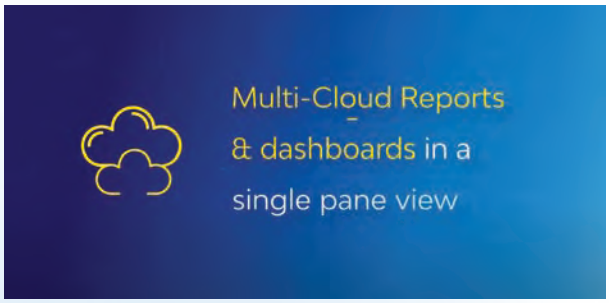
LTIMindtree is an AWS Premier Tier Services Partner and a long-standing member of the AWS Well-Architected Partner Program. The consultants at LTIMindtree spend thousands of hours every year helping customers build AWS-focused CCoE and adopt AWS Well-Architected best practices. LTIMindtree has been helping its customer adopt cloud computing, reinvent, and optimize for many years now and have helped many fortune 500 companies in this journey.

Infinity Ensure is a cloud governance platform designed to optimize governance, compliance, and remediation on Amazon Web Services (AWS), Azure, GCP, and other cloud portfolios. Infinity Ensure is an autonomous multi-cloud platform that empowers outcome-based transformations, providing continuous cloud health monitoring and threat governance. With over 1,500 checks, Infinity Ensure helps customers achieve optimized cloud environments. It ascertains whether your cloud portfolio aligns with best practices and derives strategic recommendations based on evaluated outcomes.

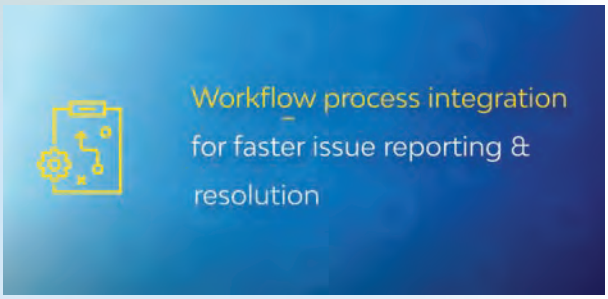
Some of the key features that Infinity Ensure offers are:

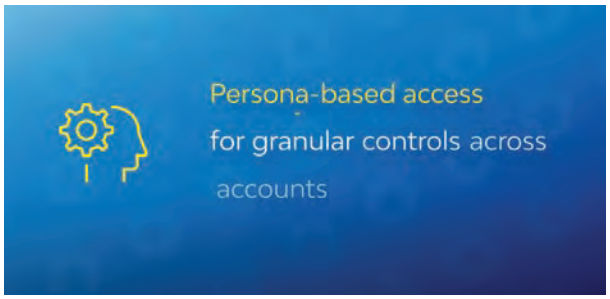
1. Automated AWS well-architected review audits
2. Advanced cost analytics
3. Unified reports and dashboards
4. Regulatory compliance checks
5. Infrastructure-as-code audits
6. Custom policies compliance checks
7. Workflow integration with ITSM tools
8. Stabilized security posture

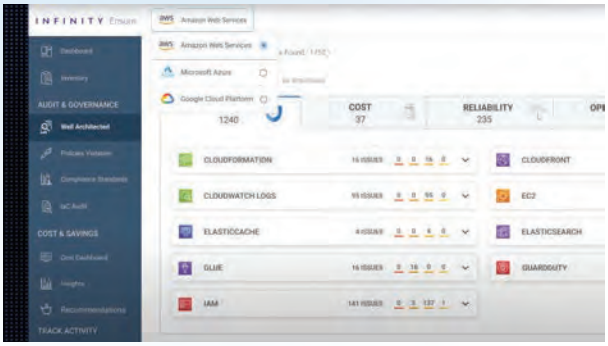
©LTIMindtree | Privileged and Confidential 2023

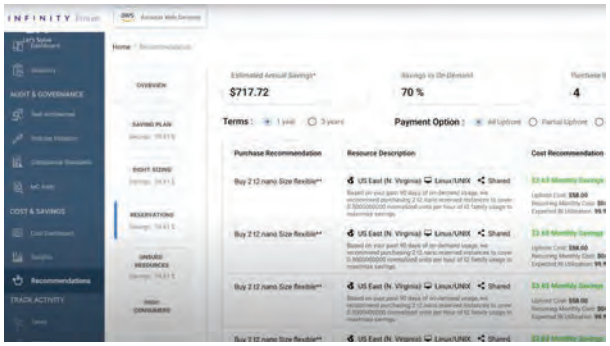


©LTIMindtree | Privileged and Confidential 2023









Refer to Strengthening Cloud Governance and Optimizing FinOps with Infinity Ensure **for more information on Infinity Ensure.**

©LTIMindtree | Privileged and Confidential 2023

About the Author

Gaurav Goel

Senior Cloud Architect

Gaurav Goel is an IT professional with 14 years of experience in cloud Architecture, cloud Consulting, DevOps, Infrastructure as Code, System Administration, and many other roles. He has worked with some of the renowned IT organizations such as LTIMindtree, HCL Technologies, IBM India, Computer Sciences Corporation, Nokia Siemens Networks, and with clients from various industries such as Telecom, Retail, Oil and Gas, Utility, Manufacturing, etc. He likes to learn new technologies and share knowledge with others.

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700+ clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by nearly 90,000 talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit **www.ltimindtree.com.**

©LTIMindtree | Privileged and Confidential 2023