When It Comes To Enterprise Cloud Economics, What You See Isn't Always What You Get

By Thomas Martin

As enterprise organizations navigate their way through the explosive growth in new technologies, tools, and services, cloud-based technologies have become the holy grail for modernizing their IT infrastructure. There are many business drivers leading to this mass migration to the cloud, but for most organizations, it is the promise of only paying for what is used when you use it. They have a mental image of eliminating the cost of the data center and paying less per hour for the same compute power in the cloud.

While this sounds like a no-brainer, you need to be smart about how you approach this migration in order to maximize your cost savings. The very nature of how cloud-based infrastructure services are created and consumed has disrupted the baseline assumptions of the traditional infrastructure leader – stable application workloads and predictable growth. And while there are many benefits to moving to the cloud, it also can create more complexity and cost for the enterprise.

For example, self-service consumption issues in the cloud can occur when several applications are being spun up without control, leading to statically run or orphaned resources and over-provisioning. You can imagine how attractive it is for creative application owners and infrastructure managers to try new services, resources, and approaches that promise to make the company more efficient. This can result in poorly architected solutions and a massive number of untagged resources that can be difficult to manage and end up costing the enterprise more in the long run.

There are three core challenges that the enterprise application transformation leader faces when moving to the cloud:

- Application selection, to-be design architecture, and resource selection
- Day 2 operations – practices and management of workloads
- Data center optimization through cloud transformation

Application selection, to-be design architecture and resource selection

Foundational to the transformation strategy is understanding your application inventory, the current technology stacks, and how these applications function within the ecosystem of the enterprise.

An example of concern would be an application that is very chatty, or has large data transfers and has many ecosystem application dependencies in the physical data center or a need for high availability across global regions. In this scenario, it may be best to leave the database instances behind in the physical data center until more of the ecosystem dependent applications are also moved to cloud-based infrastructure.

This is due to the cloud provider costs of data transfer. Using Google Cloud Platform, it is free to move data within the same location (for example us-east1), but there is a cost to move data

between multi-regions and these costs go up even further between worldwide locations. This is similar to other providers and must be taken into consideration when architecting the to-be design.

When it comes to resource selection for AWS customers, it is typical for EC2 costs to be +70% of the monthly bill. It has been my experience that when an application architect is faced with pulling an application out of a shared physical datacenter compute instance and moving to cloud-based resources, the selected compute resources are almost always initially oversized. To further compound this cost problem, the failover instances will also be oversized, and even though the team has auto-scaling set up and properly configured; if the compute resource never reaches triggering load you aren't taking advantage of the true benefits of cloud-based services.

Day 2 operations – Practices and management of workloads

Unfortunately, as indicated, most deployed compute instances are oversized; even worse, many organizations don't set up development and test resources to be shut down or suspended during non-working hours. Treating cloud-based resources with standard datacenter operations is a big flaw early in the maturity curve of newly formed DevOps teams. This also includes taking into account considerations related to releasing AWS elastic IPs when instances are stopped.

Proper tagging/labeling of these compute resources and establishing policies that are enforced to manage when non-critical resources are stopped, and the frequency of how often and how long snapshots are taken and retained are key parts of managing cloud-related costs.

"Optimizing during cloud transformation is similar to playing the game of Tetris"

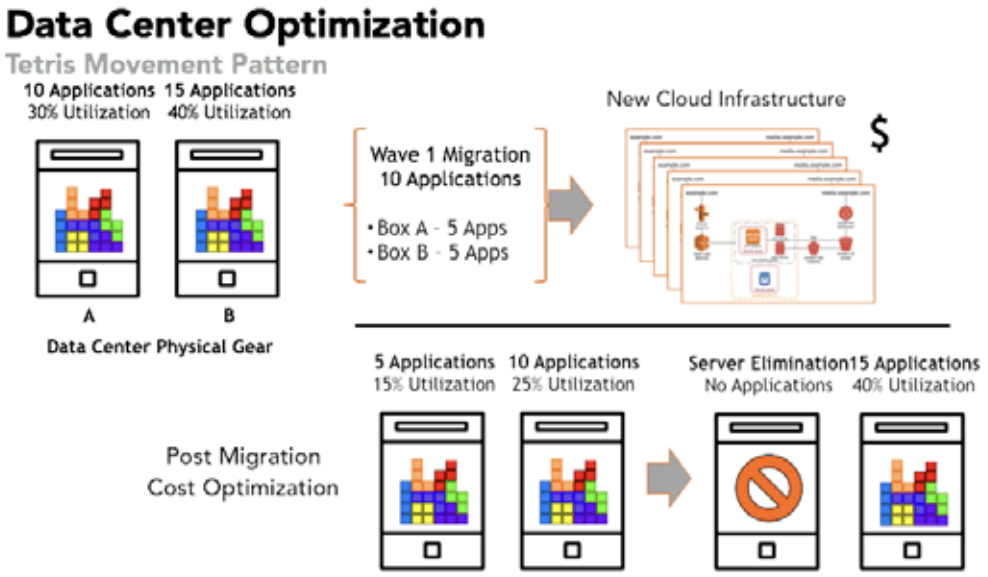
Datacenter optimization through cloud transformation

Another challenge is enterprise data center utilization. The industry average for data center utilization is around 30 percent, but it really should be closer to 70 percent or more. This is like owning a car and paying for insurance, maintenance, gas, etc., but only using it 30 percent of the year. It results in a big waste of money.

The primary reason for data center underutilization is under-optimized resources. Leased/purchased servers are running 24/7 and are sized by peak application workloads, but these workloads have high variability and many periods of inactivity.

Optimizing during cloud transformation is similar to playing the game of Tetris. As you may recall, the object of Tetris is to manipulate the squares to create a horizontal line with no gaps. In order to eliminate the cost of abandoned resources in the data center post-migration, you need to move applications in the same way. Most large companies have anywhere between 3,000 to 5,000 applications that can be either eliminated, consolidated or moved to the public cloud.

The key is to shift workloads, eliminate abandoned resources, and establish guardrails to monitor and control your cloud infrastructure.



If you only move five applications to the cloud and are paying for a box that holds ten, you are still paying for the power in the data center plus the new cloud infrastructure, etc., and the resources to manage both. Instead of saving money, you've only diluted your computing power and are actually paying for more infrastructure. Some companies end up paying double or more while they are moving from physical data centers and the cloud.

Bringing it all together

In order to truly benefit from a shift to cloud, both financially and resource-wise, you have to optimize your IT infrastructure across both the data center and the cloud. This can be accomplished by retooling and automating many of the tasks involved.

Here are five strategies for success:

1. Create an application inventory: Identify all existing applications, their function within the ecosystem, and assign unique identifiers to each one.
2. Classify your data: Set up tiers for your data based on levels of risk – from basic information to more proprietary and confidential information that needs to have a higher level of protection.
3. Create an official tagging strategy and enforcement policy: Tag your assets and utilize categorized and time-based resource use policies to control costs.
4. Actively monitor your cloud infrastructure: Cloud automation tools should be implemented to continuously monitor activity and automatically take action to correct problems.
5. Utilize an autonomous compliance engine: Establish tolerance limits and take action.

Although a cloud-first approach does offer cost savings, in order to realize those savings, you have to retrain your user base as you mature to prevent the inevitable runaway costs. You should ask yourself - can I run this workload more efficiently in the cloud? And am I using the cloud in a way that doesn't result in adding more costs the balance sheet? Having a strategy and methodology around the usage of the cloud will help you effectively manage the cost of cloud.

Thomas Martin

Thomas Martin is a former CIO at General Electric and is the founder of BigCo. SmallCo. Thomas is a transformational IT Executive with a broad functional business background. He has over 25 years of executive-level experience in the highly regulated industrial manufacturing and service sectors of Aerospace, Nuclear, Gas Turbine Energy, and Oil & Gas.