

Whitepaper

Azure Cost Management and Optimization using Tech Mahindra's FinOps Service

1



Abstract

Providers of cloud computing services aim to maximize their revenues while their customers aim to minimize their costs. We will focus on consumer's perspectives in this paper. The services provided by cloud computing are divided into software application services, platform services, and infrastructure services. Providers often bundle their software services with infrastructure, and platform services as part of a bundle. Platform services are also offered as a bundle with infrastructure services by providers.

The bundling of services prevents customers from buying software from one provider and the underlying platform or infrastructure from another. As a result of this bundling policy, consumers will be locked into specific service providers and an unfair pricing model will develop. This paper provides examples of different cost optimization levers organizations can use in the real world to optimize their spiraling cloud spend in Azure.

Key Takeaways

- ▶ Gain visibility into cloud cost
- ▶ Take back control of spiralling cloud spend
- ▶ Enforcing reduction of cloud spend
- ▶ Right-size compute and storage resources to increase efficiency
- ▶ Leveraging commercial/licensing constructs of Azure to reduce cost
- ▶ Reduce overall Azure spend by 10-20%
- ▶ Follow well-architected framework
- ▶ Right-sizing of virtual machines
- ▶ Remove/cleanup orphaned disks
- ▶ Remove/cleanup unattached public IPs
- ▶ Underutilized VM identification
- ▶ Auto shutdown for virtual machines
- ▶ Shutdown and decommission resources not used in 90 days
- ▶ Right-sizing of redundancy
- ▶ Utilizing both vertical and horizontal scaling (in and out)
- ▶ Candidates for reserved instance
- ▶ Azure hybrid benefit savings
- ▶ Monthly budget deviations
- ▶ Azure advisor recommendations for cost optimizations
- ▶ Operationalizing the IT estate in cloud



Introduction

The purpose of this white paper is to provide best practices for optimizing cloud cost, so that organizations can maximize their return on investment (ROI) as they move their workloads to the cloud. A move to the cloud enables global innovation, speeds up feature velocity for faster time to market, and increases customer satisfaction by enabling quick response to problems. It is also a good idea to audit your existing cloud applications to ensure you are following best practices. Utilizing cloud resources in the most effective and efficient manner maximizes business value while optimizing costs.

The scalability, availability, security, and pay-as-you-go features are sometimes overcompensated in the cloud. Therefore, it is important to understand how to build a cloud model for your business that will account for every single dollar.

The client is one of the global leaders in the oilfield services industry. Oil and gas producers throughout the world rely on the client for services, tools, equipment, and an exceptionally trained workforce. There was no clue of what was happening in the customer's environment as Azure consumption grew rapidly. After Tech Mahindra implemented FinOps, the customer's Azure spend is now more manageable, and the customer eventually could save 20% on their monthly Azure spend.

Industry Landscape

According to a study by a leading analyst firm, enterprises believe that more than 30% of their cloud spending is being underutilized. A major challenge for IT leaders is to predict how workloads and fluctuations will evolve, as well as the way in which new applications will consume cloud services. As a result, IT managers often fail to adequately plan when subscribing to cloud services.¹

A new paradigm for financial management is being developed called FinOps (short for cloud financial management) to address cost management challenges in cloud.

FinOps is an evolving, "FinOps is an evolving cloud financial management discipline and cultural practice that enables organizations to get maximum business value by helping engineering, finance, technology and business teams to collaborate on data-driven spending decisions."

A key aspect of FinOps is that it is a cultural practice. A best-practices group gives teams guidance on cloud usage, while everyone takes responsibility for their cloud costs. As a result, cross-functional teams from engineering, finance, and product work together to improve financial transparency and control.

With cloud computing becoming the norm in the industry, Tech Mahindra has designed and built FinOps as a service to help customers across the globe optimize their cloud spend and capitalize the savings on business growth.

Tech Mahindra's FinOps service helps business to reduce their cloud spend by:



Right-costing for the resources – PayG vs RI, BYOL vs cloud licenses



Analyzing metrics values defined to identify anomalies (For example, sudden spike or crash in usage) and trigger alerts to investigate



Right-sizing services to an optimal configuration based on the usage trend



Implement chargeback (or showback) to ensure accountability on the spend across departments/divisions



Planning right architectural design considering the business needs



Disposing unused resources



Forecast the usage based on appropriate assumptions



Analyzing usage trends and create auto-shutdown/ auto-start schedules



Establishing budgets for each application based on expected/past trends



Upgrading instance generation to deliver better performance at lower/same costs



Defining meaningful tags and enforce them for all resources in the cloud for proper cost governance



Establishing a DevOps feedback loop for cost reduction



Defining tracking metrics (for example: actual vs budget, usage trends) and build dashboard to identify optimization opportunities

Choosing Virtual Machine Payment Methods

Azure IaaS is a consumption-based service, so efficient utilization of the cloud resources can significantly reduce costs.

Virtual machines (VMs) are the single biggest cost component of a typical IT environment in Azure. We will therefore begin by reducing this large consumption component. There can be significant savings.

There are several ways to pay for Azure VMs and other cloud resources, including pay as you go, reserved instances, and spot instances.

Pay As you Go

Pay as You Go pricing model is typically suitable for users who prefer flexibility or prefer to convert capital expenses into operating expenses, as well as those who have volatile or short-term workloads or instances that cannot be interrupted.

Using the PAYG model, VMs are billed per second and can start or stop the service at any time.

Schedule VM Start/Stop

Most of the time, we do not need the compute resources during off-hours or weekends and keeping the VMs running during off-hours can increase costs. Manually shutting down an Azure VM to put it in the Stopped (deallocated) status can save you a lot of money. However, you need to remember to stop the VM when you are done.

A solution for automatic start-up/shutdown can be implemented by using an [Azure Automation account](#).

Snooze Environment

Pausing bills in Azure is not limited to the VMs. Refer to the following articles to pause billing or snooze environment when not in use:

- ▶ [Configure auto shutdown for labs and VMs in DevTest Labs](#) - DevTest Labs
- ▶ [Automatically start lab VMs with auto-start in Azure DevTest Labs](#) - DevTest Labs
- ▶ [Pause and resume compute in dedicated SQL pool](#) - Azure Synapse Analytics
- ▶ [Use autoscaling feature to dynamically increase/decrease compute resources](#) - Virtual Machine Scale Set, Azure App Service

Pay As you Go

Reservations helps organizations using Azure cloud to save money by committing to a usage plan for multiple products (For example: virtual machines, storage, Azure SQL database, Azure SQL managed instance). Committing allows you to get a discount for resources you are using. Reserved Instances are one of the costs saving strategies that can give you the biggest saving. The percentage of savings may vary based on Azure region, instance type and commitment to one or three-year benefit.

Azure Reservations are ideal for workloads with a consistent resource usage over a period of 1 or 3 years, whereas PAYG will be suitable for variable workloads.

By reserving virtual machines (VMs) on Microsoft Windows and Linux for a year or three years, one can save up to 72 percent in comparison with PAYG prices. By combining Azure RIs with Azure Hybrid Benefit, one can save up to 80%.

With reserved instance size flexibility, you can further reduce costs.

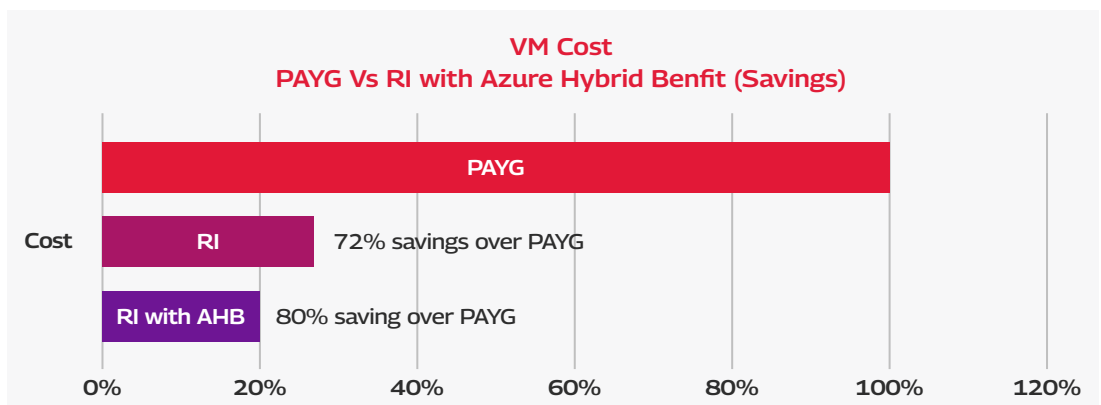


Figure 1 - Compare VM Cost -PAYG Vs RI & AHB

Additionally, you can better budget and forecast with a single upfront payment, making it easier to calculate your investments. You can also lower your upfront cash outflow with no additional fees with monthly payment options.

Spot Instance

The use of spot instances can significantly reduce your cloud costs. With spot instances, you can purchase spare computing capacity and receive substantial discounts up to 90% over PAYG pricing on Azure. These Azure spot instances can be taken away with little notice if demand for capacity increases or instances are needed to service reserved instances or pay-as-you-go customers.

One can significantly reduce the cost of running your workload in Azure by using Spot virtual machines if your workload can tolerate interruptions and its execution time is flexible. Use Azure VMs or VM scale sets to run your workloads.

Spot virtual machine has the following advantages:

- ▶ Spot virtual machines allow you to save up to 90% over pay-as-you-go
- ▶ Low-priority workloads can be run on spot VMs, and a lot of costs can be saved
- ▶ A maximum price can be set for spot virtual machines depending on what you are willing to pay

The following types of workloads are ideal for spot VMs:

- ▶ Dev/test environments, including continuous integration and continuous delivery (CI/CD) workloads
- ▶ Select high-performance computing scenarios, batch processing jobs, or visual rendering applications
- ▶ Big data, analytics, container-based, and large-scale stateless applications

Azure Hybrid Benefit

A cost saving benefit that lets you bring your existing on-premises Windows Server and SQL Server licenses with active Software Assurance or subscriptions to Azure with no additional cost. Opportunity to save up to 80% (actual savings may vary based on region, instance type or usage) compared to PAYG rates and achieve the lowest cost of ownership when it's combined with Reserved Instance. Applies to Windows Server, SQL Server, RedHat and SUSE Linux subscriptions.

Use the Azure Hybrid Benefit Savings Calculator to calculate your Windows Server, SQL Server, SQL Database and SQL Managed Instances cost and realize AHB savings Vs PAYG license cost.

Right-Sizing

Right-Sizing Virtual Machines

Most companies that are successful at avoiding waste are optimizing resource usage in a decentralized fashion, as individual application owners are usually the best equipped to shut down or resize resources due to their in-depth familiarity with the workloads.

In addition, you can use Azure Advisor to help detect issues like overprovisioned VM instances or idle resources. Enabling your team to surface these recommendations automatically is the aim of any great optimization effort.

VM Metric	Description
Available Memory Bytes	Amount of physical memory, in bytes, immediately available for allocation to a process or for system use in the Virtual Machine
Data Disk Bandwidth Consumed Percentage	Data disk bandwidth consumed per minute as a percentage
Data Disk IOPS Consumed Percentage	Percentage of data disk I/Os consumed per minute
Data Disk Read Operations/Sec	During monitoring, read IOPS from one disk
Data Disk Write Operations/Sec	Write IOPS from a single disk during monitoring period
Network In Total	The number of bytes received on all network interfaces by the Virtual Machine(s) (Incoming Traffic)
Network Out Total	The number of bytes out on all network interfaces by the Virtual Machine(s) (Outgoing Traffic)
OS Disk Bandwidth Consumed Percentage	Operating system disk bandwidth consumption per minute as a percentage
OS Disk IOPS Consumed Percentage	Percentage of operating system disk I/Os consumed per minute
OS Disk Read Operations/Sec	During monitoring period for OS disk, read IOPS from a single disk
OS Disk Write Operations/Sec	Write IOPS from a single disk during monitoring period for OS disk
Outbound Flows	Outbound Flows are number of current flows in the outbound direction (traffic going out of the VM)
Percentage CPU	Virtual Machine(s) utilizing the percentage of compute units allocated to them
VM Cached Bandwidth Consumed Percentage	The percentage of cached disk bandwidth consumed by the virtual machine
VM Cached IOPS Consumed Percentage	Percentage of cached disk IOPS consumed by the VM
VM Uncached Bandwidth Consumed Percentage	Percentage of uncached disk bandwidth consumed by the VM
VM Uncached IOPS Consumed Percentage	Percentage of uncached disk IOPS consumed by the VM

Figure 2 - VM Metrics in Azure Monitor for Right-Sizing

Following decision tree has been used to resize VM with the intent of CPU and memory utilization to be 80% on average or lower while leaving 20% as buffer for spike utilization:

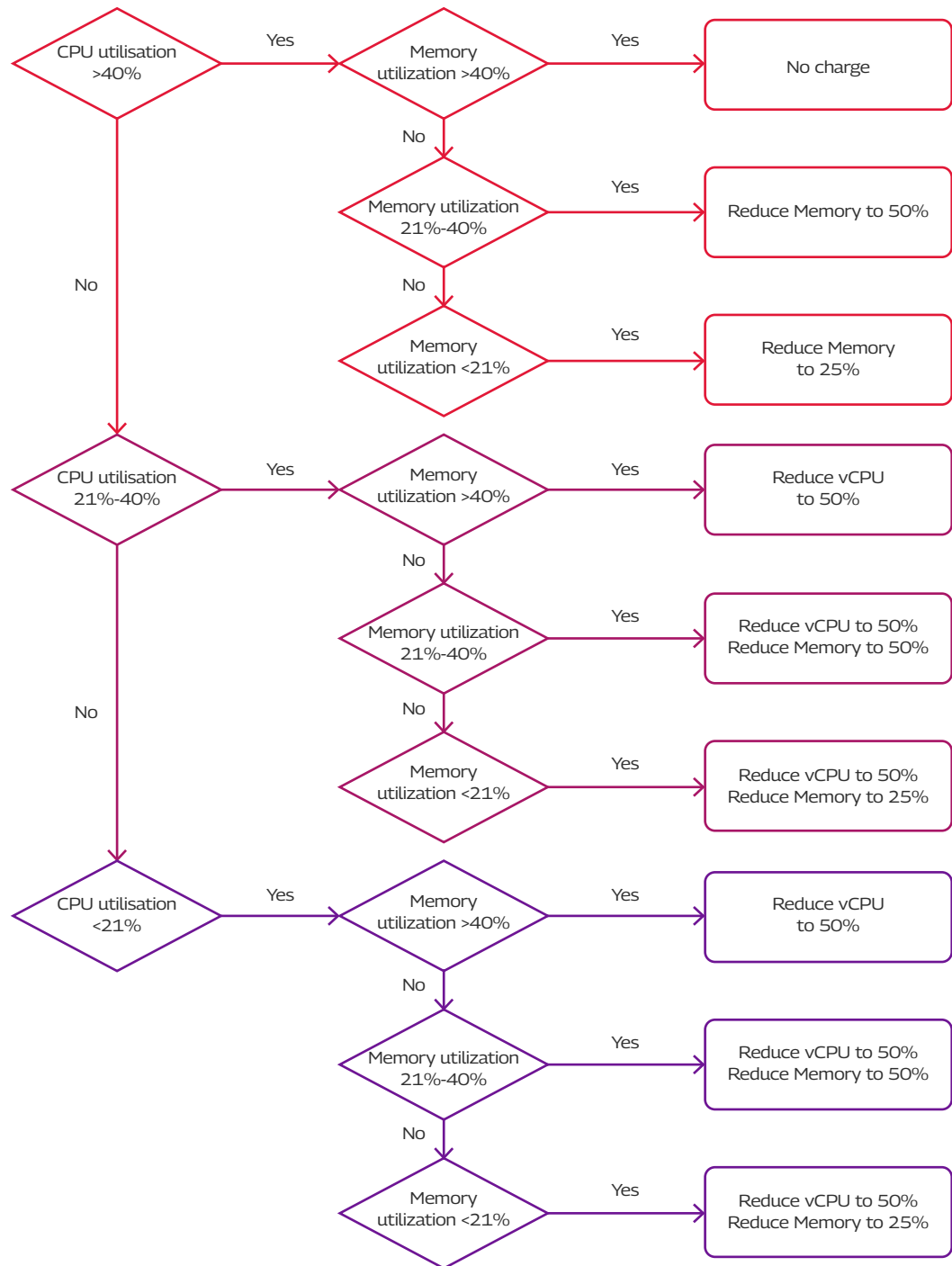


Figure 3 - VM Resizing Decision Tree

Following exemptions are given for VMs with special needs:

- ▶ VMs configured with large number of data disks
- ▶ VMs IOPS demands is much higher, and it cannot be processed by small size VM
- ▶ Virtual appliances such as Firewall, SD-WAN etc.

Comparing VM Reserved Instance Cost (3 Years Term) - Current Vs Proposed Sizes:

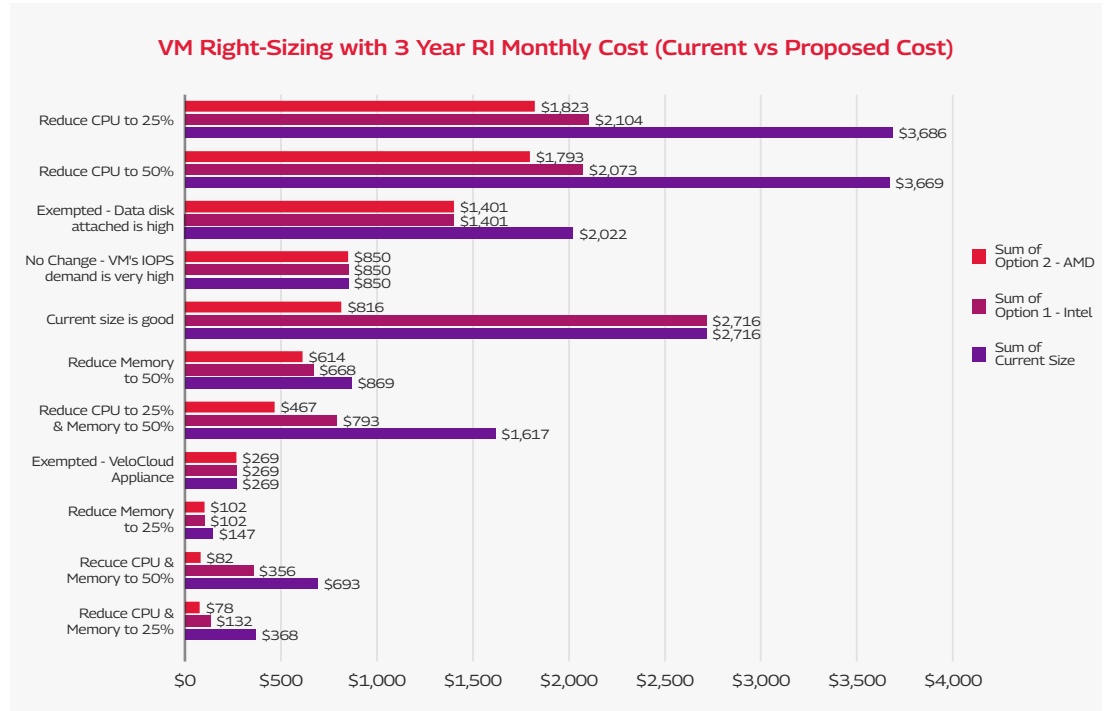


Figure 4 - VM Resizing Cost Comparison

With the VM right-sizing in place, the customer can save approximately 43% and 63% on Intel and AMD CPU respectively which contributes to a savings of ~4% on their spiralling Azure spend.

Right-Sizing Managed Disks

We collected Azure Monitor metrics available for the managed disks used in the environment and analysed IOPS of Read and Write operations for last 90 days to come up with an average during business hours and spike in disk operations to arrive at a value which will be optimal for the application.

In Azure Monitor, the following metrics were used to get insight into Disk IOPS, throughput, and queue depth on the VM level:

- ▶ **OS Disk Queue Depth:** Current outstanding IO requests waiting to be read from or written to the OS disk.
- ▶ **OS Disk Read Operations/Sec:** Reads from the OS disk in a second.
- ▶ **OS Disk Write Operations/Sec:** Writes from the OS disk in a second.
- ▶ **Data Disk Queue Depth:** The number of IO requests waiting to be read from or written to the data disk(s).
- ▶ **Data Disk Read Operations/Sec:** Reads from a data disk(s) in a second.
- ▶ **Data Disk Write Operations/Sec:** Writes from data disk(s) in a second.

You can use the following metrics to diagnose bottlenecks in your virtual machine and disk combination. The following configuration is required to access these metrics:

- ▶ This feature is only available for premium storage enabled VM series
- ▶ The metrics are not available for ultra-disks, but they are available for all other disk types on these VM series

Metrics that help diagnose disk IO capping are:

- ▶ **Data Disk IOPS Consumed Percentage:** Percentage of data disk IOPS completed over provisioned data disk IOPS. When this amount is at 100%, your application's IOPS are capped by the data disk's limit.
- ▶ **Data Disk Bandwidth Consumed Percentage:** It is calculated as the percentage of completed data disk throughput over provisioned data disk throughput. Your application is IO capped at 100% if this amount is at 100% of your data disk's bandwidth limit.
- ▶ **OS Disk IOPS Consumed Percentage:** Percentage of OS disk IOPS completed over provisioned data disk IOPS. When this amount is at 100%, your application's IOPS are capped by the OS disk's limit.
- ▶ **OS Disk Bandwidth Consumed Percentage:** It is calculated as the percentage of completed data disk throughput over provisioned OS disk throughput. Your application is IO capped at 100% if this amount is at 100% of your OS disk's bandwidth limit.

The following decision tree has been used to resize managed disks attached the VMs:

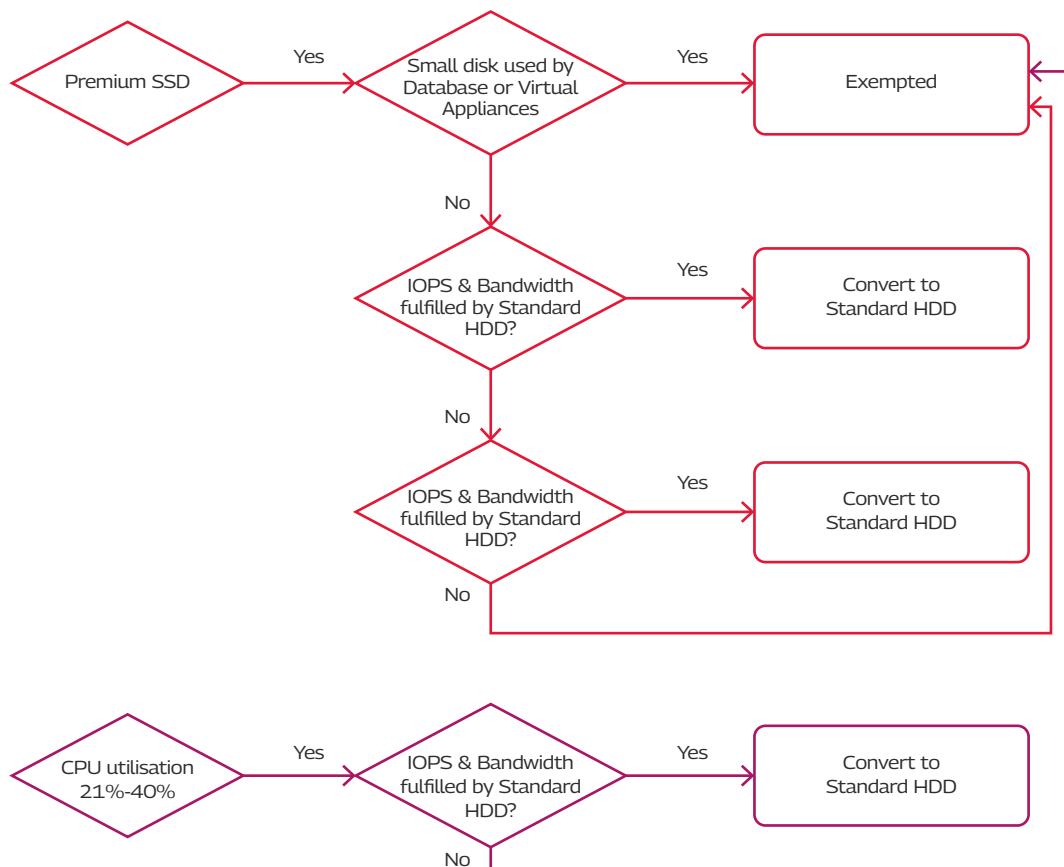


Figure5 - Managed Disk Resizing Decision Tree

Note: It's extremely important to review the proposal with the application owners and make necessary exemption/amendments as they are the best person to advise on obtaining optimal performance from the application standpoint.

Comparing Managed Disks Cost - Current Vs Proposed size:

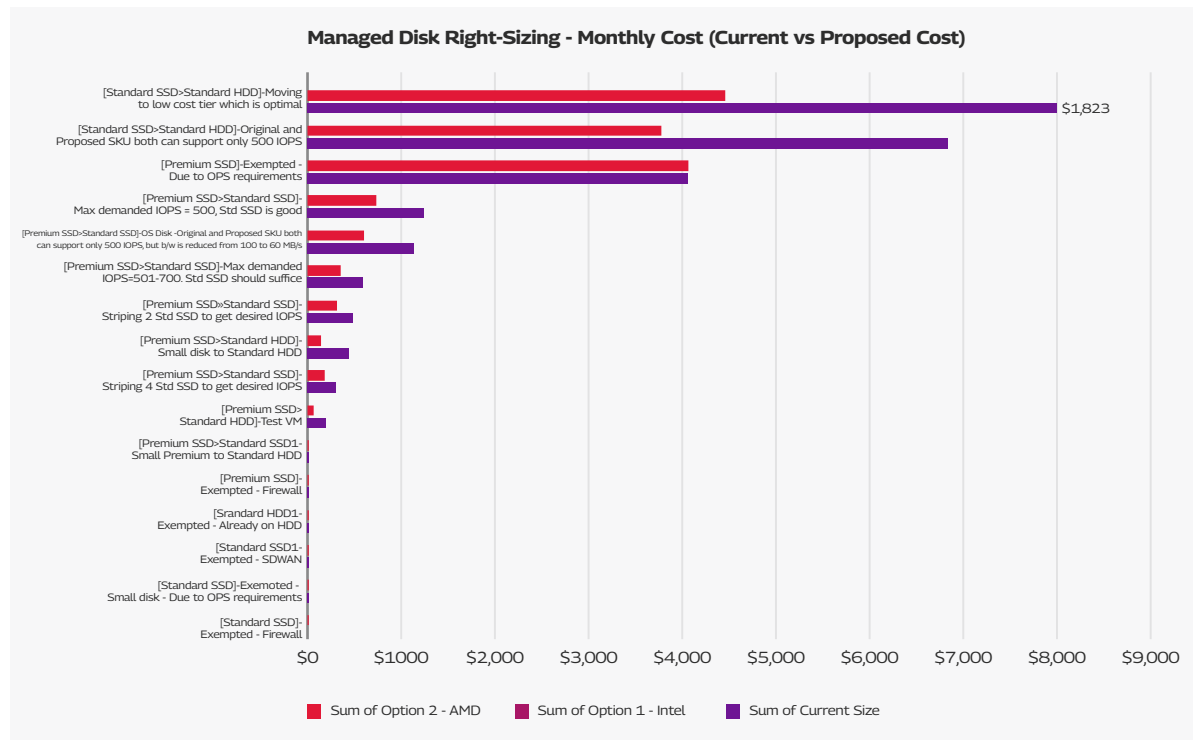


Figure 6 - Managed Disks Resizing Cost Comparison

With the managed disk right-sizing in place, the customer can save approximately 37% on managed disk which contributes to a savings of ~8% on their spiralling Azure spend.

Right-Sizing Backup Storage

Backups are an essential part of any business continuity and disaster recovery strategy because they protect your data from corruption or deletion. Azure Backup offers three types of replications to keep your backup storage/data highly available.

- ▶ **Locally redundant storage (LRS):** Syncs data synchronously across three disks in the primary region. Provides moderate availability at a lower cost.
- ▶ **Geo-redundant storage (GRS):** GRS is the default and recommended replication option. GRS replicates your data to a secondary region (hundreds of miles away from the primary location of the source data). GRS costs more than LRS, but GRS provides a higher level of durability for your data, even if there's a regional outage.
- ▶ **Zone-redundant storage (ZRS):** ZRS replicates your data in availability zones, guaranteeing data residency and resiliency in the same region. ZRS has no downtime. So, your critical workloads that require data residency, and must have no downtime, can be backed up in ZRS.

If you have data that does not need to be restored outside of your primary region and you want really a low-cost backup, then LRS may be the choice.

Product Name	Meter Description	Billable Quantity	TermAnd BillingCycle	Backup Storage Cost (USD/Monthly)	
				GRS	IRS
Backup	Backup - GRS Data Stored (GB/Month)	365,755	GRS Data Stored (GB/Month)	\$13,928	\$6,964
Backup	Backup - GRS Data Stored (GB/Month) - «Region1»	912	GRS Data Stored (GB/Month)	\$46	\$23
Backup	Backup - GRS Data Stored (GB/Month) - «Region2»	5,339	GRS Data Stored (GB/Month)	\$254	\$127
Total Cost				\$14,228	\$7,114
Savings (USD/Monthly)				50%	

Figure 7 - Comparing Backup Storage Redundancy Cost - GRS vs LRS

The customer can save ~50% on the backup storage cost by moving from GRS to LRS on selective non-production backup data which contributes to a savings of 6% on their spiralling Azure spend.

Delete Unused and Unnecessary Resources

Delete Unused Virtual Machines Disks

Even services which can be stopped and deallocated such as Virtual Machines will cost you money if they are not deleted. If you don't delete unused instances in Azure, you still pay for all the associated costs including the storage which can be significant. Moreover, you need to specifically delete your storage when you delete those VMs or you will keep paying for that too. That is known as orphaned storage or unattached storage, and it is not cheap.

Cost of Unattached Disk Storage:

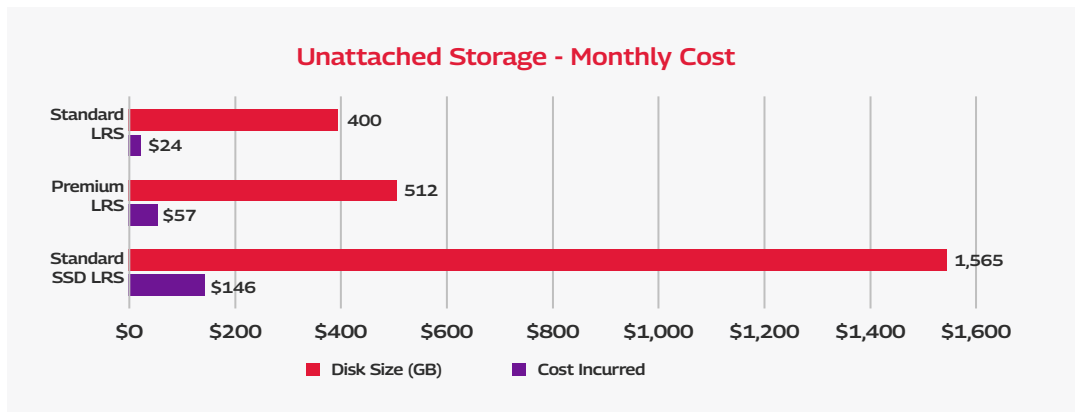


Figure 8 - Saving from Unattached Storage

By eliminating unattached storage, the customer can save ~1% on their spiralling Azure spend.

Delete Unassociated Network Interface Cards

When you delete a VM, the network interface cards (NICs) are not deleted by default. If you create and delete multiple VMs, the unused NICs continue to use the internal IP address leases. As you create other VM NICs, they may be unable to obtain an IP lease in the address space of the subnet. So, it is a good idea to discover unattached network interfaces and delete them to free up IP addresses.

Delete Unassociated Public IP Addresses to Save Money

Using Azure Advisor, you can identify public IP addresses that are not associated with Azure resources such as load balancers or virtual machines. There is a nominal charge associated with these public IP addresses. You can save money by deleting them if you don't intend to use them.

SQL Managed Instance reserved capacity

Azure Advisor analyses SQL Database and SQL Managed Instance usage patterns over the past 30 days. It then recommends reserved capacity purchases that optimize costs. By using reserved capacity, you can pre-purchase SQL DB hourly usage and save over your SQL compute costs. Your SQL license is charged separately and isn't discounted by the reservation. Reserved capacity is a billing benefit and automatically applies to new and existing deployments. Advisor calculates savings estimates for individual subscriptions by using 3-year reservation pricing and by extrapolating the usage patterns observed over the past 30 days.

Comparing SQL Managed Instance Cost - PAYG Vs Reserved Instance (3 Years Term):

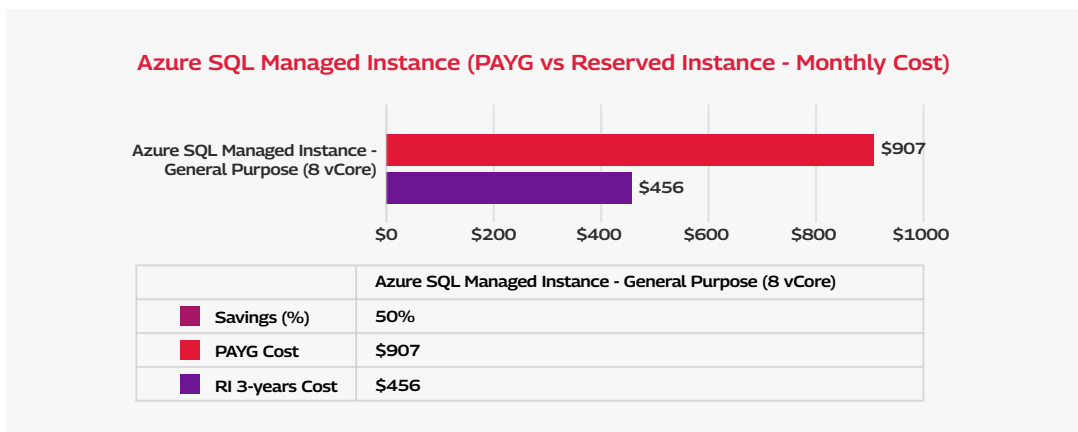


Figure 9- Comparing SQL Managed Instance Cost - PAYG Vs RI

By converting Azure SQL Managed Instance from PayG to Reserved Instance with a 3-year term, the customer can save ~50% on their spiralling Azure spend.

Conclusion

As a result of our FinOps consulting, we have helped the customer to save approximately 20% on their spiralling Azure spend in the first phase of cost optimization.

Tech Mahindra's FinOps is a strategic and tactical service for Azure cloud cost management and optimization. With FinOps, we have helped organizations across the world to take back control of spiralling cloud spend and enforcing the reduction of cloud cost with our unique methodologies and tools. We are cloud agnostic and our FinOps framework can be used across Azure, AWS, and Google Cloud Platform. We use a combination of cloud native tools and third-party tools such as NetApp Spot, CloudHealth, ParkMyCloud, and FinOptimo to optimize spiralling cloud spend.

Cloud cost optimization is important because you want to maximize your business benefits by optimizing your cloud spend. Cloud cost optimization is the process of reducing your overall cloud spend by right-sizing cloud services to scale, identifying overprovisioned resources, improve resource efficiency, right-costing by eliminating waste and reserving capacity for higher discounts.

Tech Mahindra is positioned as a leader by notable analysts in the cloud space. We have successfully implemented large scale Azure transformational deals using our Agile-based migration delivery methodology resulting in 3X faster business adoption.

Authors



Guru Prasad C P

Group Practice Head, Azure Cloud Services

Guru Prasad C P has an experience of over 22 years with over 8 years specifically in the public cloud working in Asia, ANZ, Europe and the US. His experience includes, setting up practice teams aligned to industry verticals and horizontals, analyst interactions for positioning the offerings, hiring the right talent, involving in strategic exercise mergers and acquisitions, organization building, and creating frameworks and IP's. At Tech Mahindra he is responsible for practice and competency development which includes alignment with OEMs for solutions, offerings and adoption of new technologies, customer interfacing where he acts as a trusted advisor in providing unbiased views/opinions and aligning with organization goals at the same time, value creation, developing practice areas deal making, solution support for large deals, and carve out deals from azure and hybrid cloud perspective



Mahendra Singh

Principal Solution Architect, Practice Lead (Cloud Infrastructure and Migration)

Azure Infra and networking @ Tech Mahindra's Microsoft Azure Competency With over 20 years of experience, Mahendra is a core cloud Architect with a Business Consultant's Mindset for technology transformation & modernization across all divisions and with Horizontal Focus on Cloud Landscapes. His experience and cloud qualifications blend well to provide advisory services to the clients on best suited direction into the public cloud brands with hybrid approach. He is certified with Azure, AWS, Oracle and Google Cloud brands. He has been instrumental in driving cloud adoption and migration strategies across various clients on multiple cloud platforms. Mahendra leads the division of cloud pre-sales in the area of infrastructure transformation across various geographies.