# 9 Best Practices for Reducing Spend in Your Multi-Cloud Environment

Get Started

# Introduction

Many organizations have learned the hard way that moving to the public cloud doesn't always yield the cost savings expected.

This doesn't mean that moving to the public cloud is a mistake. The public cloud provides huge benefits in agility, responsiveness, simplified operations, and improved innovation. The mistake is assuming that migrating to the public cloud without implementing optimization and governance practices will lead to cost savings.


As a result, many organizations take a strategic approach by having a multi-cloud environment to take advantage of the best services and pricing discounts that each cloud provider has to offer. While this approach presents numerous cost-saving opportunities, the challenge becomes ensuring all of your cloud environments are properly managed and under control.

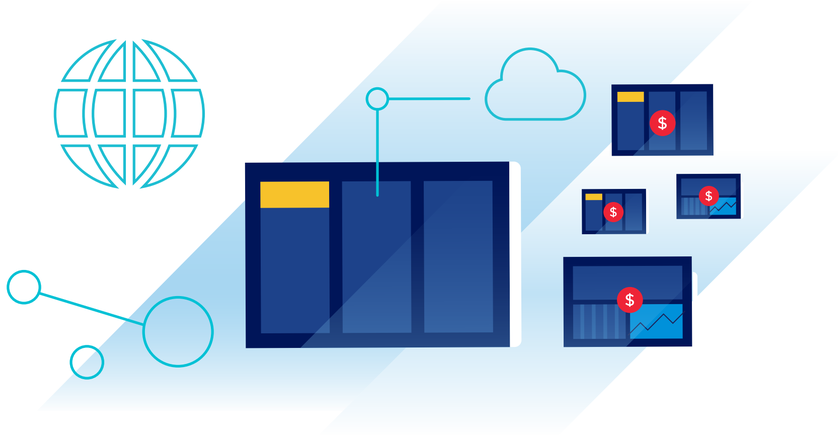
The first step in combating rising cloud costs is to gain visibility across your organization's entire multi-cloud spend. Once you've identified the areas of high and/or rapidly growing costs, use these proven best practices for cost reduction and optimization to make sure you get the most out of your cloud investment.

# 1. Delete unattached virtual server disk infrastructure

It's common to see thousands of dollars in unattached virtual server disk infrastructure being spent within your cloud accounts. Each cloud provider has a different name for virtual server disk infrastructure, such as Amazon Elastic Block Store (EBS) volumes, Azure Disk Storage, and Google Persistent Disk. Generally, these virtual server disks are costing money but aren't being used for anything.

For example, when an Amazon Elastic Compute Cloud (EC2) instance is launched, an EBS volume is usually attached to act as the local block storage for the application. When the instance is terminated, it's possible that the unattached EBS volume will be left running. Amazon Web Services (AWS) will continue to charge for the full list price of the EBS volume despite the fact that it is no longer in use.

Because of the dynamic nature of cloud computing, it's easy for users to quickly spin up and spin down workloads, but that means the risk of leaving behind unattached storage is high. By continuously checking for unattached virtual server disk infrastructure, you can cut thousands of dollars from your monthly bill.



**Pro tip**

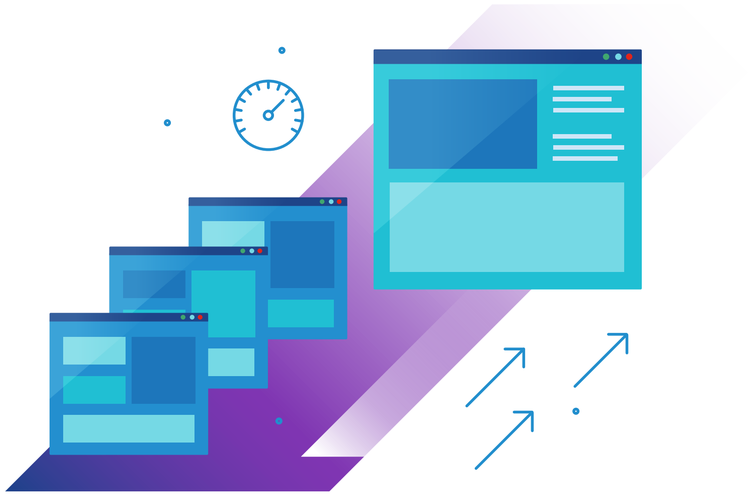
Delete virtual server disk infrastructure when it has been unattached for two weeks as it is unlikely the same storage will be utilized again.

# 2. Delete aged snapshots

Many organizations use snapshots for point-in-time recovery in case of data loss or disaster. However, the cost of snapshots can quickly get out of control if not closely monitored. Individual snapshots are not costly, but the cost can grow quickly when several are provisioned.

A compounding factor on this issue is that users can configure settings to automatically create subsequent snapshots on a daily basis without scheduling older snapshots for deletion. Organizations can help get snapshots back under control by monitoring snapshot cost and usage per virtual server to make sure they don't spike out of control.

Set a standard in your organization for how many snapshots to retain per virtual server. Remember that most of the time, recovery will occur from the most recent snapshot.



### Pro tip

One way of finding good snapshot candidates for deletion is to identify snapshots that have no associated volumes. When a volume is deleted, it's common for the snapshot to remain in your environment.

Be careful not to delete snapshots being utilized as a volume for an instance.

# 3. Delete disassociated IP addresses

An IP address is usually associated with a virtual server and allows it to be reached via the internet. The pricing structure for an elastic IP address/external IP address is unique in that when you use it, the IP address is free of charge. However, if a virtual server is terminated and the IP address is not associated, you will be charged for the disassociated IP addresses.

Unfortunately, it's difficult to identify and manage disassociated IP addresses. This task might or might not amount to significant cost savings, but it's important to develop a habit of staying on top of wasted resources and remaining proactive in managing costs before they spike out of control.

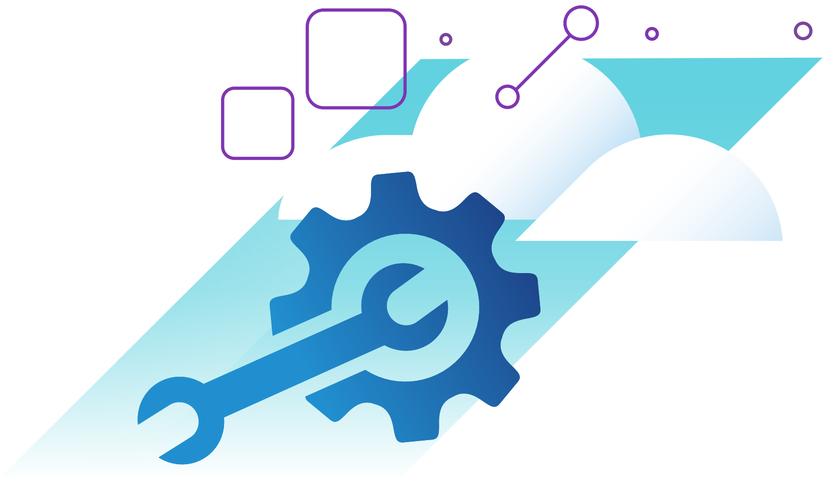
**Pro tip**

If using Google Cloud Platform (GCP), you can check whether a static external IP address is in use by making a gcloud compute addresses list request. This command returns a list of static external IP addresses and their statuses, enabling you to delete those that show a reserved status rather than an in-use status.

# 4. Terminate zombie assets

Zombie assets are infrastructure components running in your cloud environment but not being used for any purpose. Zombie assets can come in many forms—such as Amazon EC2 instances, Azure Virtual Machines (VMs), Google Compute Engine VMs, relational databases, and more—that were once used for a particular purpose but are no longer in use and have not been turned off. Zombie assets also can occur when the launch process fails or because errors in the script fail to deprovision them.

No matter the case, cloud service providers will charge you as long as these assets are in a running state. They must be isolated, evaluated and immediately terminated if deemed nonessential. Take a backup of the asset before terminating or stopping it to ensure you can recover the asset if it is needed again.



**Pro tip**

Start your zombie hunt by identifying compute services that have a max CPU of less than 5 percent over the past 30 days. This doesn't automatically mean this asset is a zombie, but it's worth investigating further.

# 5. Upgrade to the latest generation

Every few years, cloud providers release the next generation of enhancements that often offers users improved price per performance and additional functionality. For most companies, the migration from the first generation to the second generation will be a gradual process. Not all assets and services will be great candidates for conversion at the same time, especially if pricing discounts are in use, and so it's important for organizations to identify the conversions that can happen immediately and what can be postponed until a later date.

## Amazon Web Services

For AWS users, the release of next-generation instances means improved price-per-compute performance and additional functionality, such as clustering, enhanced networking, and the ability to attach new types of EBS volumes. For example, upgrading a c4e.xlarge to a c5.xlarge will result in a 25 percent improvement in price and performance compared to C4 instances.

**Pro tip**

One large SaaS company found that almost 60 percent of the instance hours they ran in the past 12 months were using older-generation instance types. Analysis revealed that upgrading those instances to the latest generation would save them millions of dollars per year.

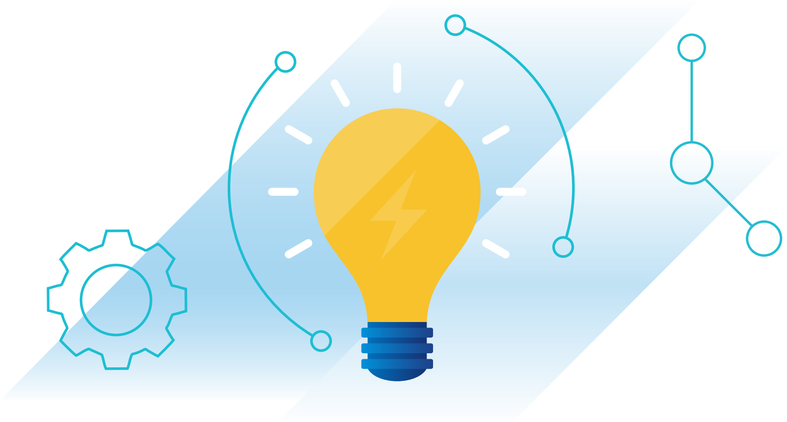
## Microsoft Azure

In 2014, Microsoft introduced the next generation of Azure deployment, called Azure Resource Manager (ARM), or sometimes v2. ARM gives you access to additional functionality, such as resource grouping, advanced tagging, role-based access control, and templates. While the prices for ARM and Azure classic (Azure v1) are the same, the management improvements can drive significant savings.

Azure users also have the ability to upgrade some VM types to the latest generation. While the new versions of Azure VMs have the same price points, they come with performance improvements that can enable you to run fewer VMs. For example, upgrading a D-series VM gives you 35 percent faster processing and greater scalability for the same price point.

## Google Cloud Platform

For GCP users, Google Cloud SQL Second Generation is the latest version for Cloud SQL for MySQL. Second Generation instances support most of the features of First Generation instances while also offering higher performance and storage capacity at a lower cost, such as increased throughput and the option to add high availability failover. Cloud SQL Second Generation determines instance pricing by the machine type, introduces per-minute billing, and allows users to take advantage of sustained use discounts.



# 6. Rightsize your environment

Rightsizing is the cost reduction initiative with the potential for the biggest impact. It's common for developers to spin up new infrastructure-as-a-service (IaaS) or platform-as-a-service (PaaS) offerings that are substantially larger than necessary. This might be intentional to give themselves extra headroom or accidental because they don't know the performance requirements of the new workload yet. Over-provisioning can lead to exponentially higher costs. Without performance monitoring or cloud management tools, it's hard to tell when assets are over- or under-provisioned.

## Compute

It's important to consider CPU, memory, disk and network in/out utilization. Reviewing these trended metrics over time, you can make decisions about reducing the size of the virtual server without hurting the performance of the application that's running on it.

## Storage

Similar to virtual servers, storage can also be rightsized. Instead of looking at CPU, memory, disk and network, the critical factors to consider with storage are capacity, IOPS and throughput. Removing unattached storage is one way to reduce spend and ensure resources aren't going unused.

## Database

Many developers use PaaS offerings, such as relational databases, to manage their applications. Therefore, it's important to evaluate how well they are being utilized in terms of the workloads you are running on them. As a best practice, you should rightsize to the lowest cost database that meets your performance requirements.

### Pro tip

A good starting place for rightsizing is to look for virtual servers that have an average CPU of less than 5 percent and a maximum CPU of less than 20 percent for 30 days. Virtual servers that fit these criteria are viable candidates for rightsizing or termination.

AWS offers several types of EBS volumes, from Cold HDDs to Provisioned IOPS SSDs, each with their own set of pricing and performance. By analyzing the read/writes on all volumes, you can find opportunities for cost savings.

Azure SQL databases can be purchased through a database transaction unit (DTU)-based model, which is a blend of compute, memory and I/O resources, or through the vCore-based model, which lets you choose the number of vCores, the amount of memory, and the amount and speed of storage. The critical factors to take into consideration include DTUs, database size, and storage.

# 7. Take advantage of commitment-based discounts

## Amazon Web Services

Amazon EC2 Reserved Instances (RIs) allow you to make a commitment to AWS to utilize specific instance types in return for a discount on your compute costs as well as a capacity reservation that guarantees your ability to run an instance of this type in the future.

Reserved Instances are purchased either all upfront, partially upfront, or no upfront and can be applied to active instances. RIs can save you up to 75 percent compared to on-demand pricing, making their use an easy decision to make for any company with sustained EC2 usage.

One common misconception about RIs is that they cannot be modified, which isn't true. Once purchased, Standard RIs can be modified in several ways at no additional cost, including switching Availability Zones within the same region and altering the instance type within the same family. Convertible RIs are also available and have greater flexibility than Standard RIs. They can be exchanged in several ways, such as exchanging for a new instance type or a new operating system.

AWS has a commitment-based pricing model known as AWS Savings Plans, which helps you save on EC2, AWS Fargate, and AWS Lambda usage.

**Pro tip**

EC2 instances aren't the only assets in AWS that have reservations. Amazon Relational Database Service (RDS), Amazon DynamoDB, Amazon Redshift, and Amazon ElastiCache also have reservations that can be purchased to help users reduce cost.

## Microsoft Azure

Azure Reserved VM Instances allow you to make a one- or three-year upfront commitment to Microsoft to utilize specific virtual machine instance types in return for discounts on compute costs and prioritized capacity. Reservations can save you up to 72 percent compared to pay-as-you-go pricing.

Similar to AWS, Microsoft also allows customers to modify reservations. These changes can include changing the scope from single subscription to shared (or vice versa), exchanging Reserved VM Instances across any region or series, and canceling your Reserved VM Instances at any time for an adjusted refund.

Microsoft allows you to achieve a greater cost savings (up to 82 percent) by leveraging Reserved VM Instances combined with Azure Hybrid Benefit. Azure Hybrid Benefit covers the cost of the Windows OS on up to two virtual machines per license, so you only have to pay for the base compute costs.

**Pro tip**

Microsoft offers reservations for Azure Virtual Machines and also allows users to purchase reservations for SQL Database compute capacity and Azure Cosmos DB throughout.

## Google Cloud Platform

Google offers users the ability to purchase a specific amount of compute or memory for discounts at no upfront cost. With a commitment of either one or three years, customers can save up to 57 percent of the normal price of cloud usage with Google committed use discounts.

These discounts can be used for standard, high-memory, high-CPU, customer machine types, and sole-tenant node groups. When they expire, Google Compute Engine VMs get charged at the normal price. It's important to note that once commitment discounts are purchased, customers cannot cancel them.

Even if you don't make the commitment to purchase a committed use discount, you can still benefit from discounts for prolonged usage. Google sustained use discounts are given to users when they consume certain resources for the better part of a billing month and are applicable to resources such as custom machines, sole-tenant nodes, GPU devices, and more. These discounts are given automatically and customers don't have to do additional work to take advantage of them.

# 8. Stop and start virtual servers on a schedule

Cloud service providers will bill for a virtual server (e.g., EC2 instance, virtual machine, etc.) as long as it's running. Inversely, if a virtual server is in a stopped state, there is no charge associated with it.

For virtual servers running 24x7, cloud service providers will bill for 672 to 744 hours per virtual server per month, depending on the month. If a virtual server is turned off between 5 PM and 9 AM on weekdays and stopped weekends and holidays, then the total billable hours per month ranges from 152 to 184 hours per virtual server per hour per month, saving you 488 to 592 instance hours per month.

This is an extreme example as having flexible workweeks and global teams means that you can't just power down virtual servers outside of normal working hours. However, outside of production, you'll likely find many virtual servers that do not need to truly run 24x7x365.

The most cost-efficient environments dynamically stop and start Amazon EC2 instances, Azure VMs, and Google Compute Engine VMs based on a set schedule. These types of lights-on/lights-off policies can often be even more cost-effective than purchasing pricing discounts, so it's crucial to analyze where this type of policy can be implemented.

**Pro tip**

Set a target for weekly hours that non-production systems should run. One large publishing company set that target at less than 80 hours per week, which saves them thousands of dollars per month.

# 9. Move object storage to lower cost tiers

Each cloud provider offers several tiers of object storage at different price points and performance levels. The best practice is to move data between the tiers of storage depending on its usage. Here's how that might look depending on which cloud provider you're working with.

## Amazon Web Services

Many AWS users tend to favor Amazon Simple Storage Service (S3), but you can save more than 75 percent by migrating older data to lower storage tiers. For example, Infrequent Access storage is ideal for long-term storage, backups, and disaster recovery content, while Glacier is best suited for data archiving.

In addition, the Infrequent Access storage class is set at the object level and can exist in the same bucket as Standard. The conversion is as simple as editing the properties of the content within the bucket or creating a lifecycle conversion policy that automatically transitions S3 objects between storage classes.

## Microsoft Azure

For Azure storage, you can adjust redundancy (how many copies are stored across how many locations) and the access tier (how often data is accessed). Microsoft allows customers to mix and match across four redundancy options and three access tier options to create the right solution.

For example, cold locally redundant storage (LRS) is ideal for long-term storage, backups, and disaster recovery content, while cold geographically redundant storage (GRS) is best suited for data archiving.

## Google Cloud Platform

Google offers four object storage classes: multi-regional (e.g., workloads with global needs such as gaming and mobile applications), regional (e.g., data analytics for Compute Engine VMs across a region), nearline (e.g., frequent usage once a month for backup and long-tail multimedia content), and coldline (e.g., infrequent usage once every year for disaster recovery or data archiving).

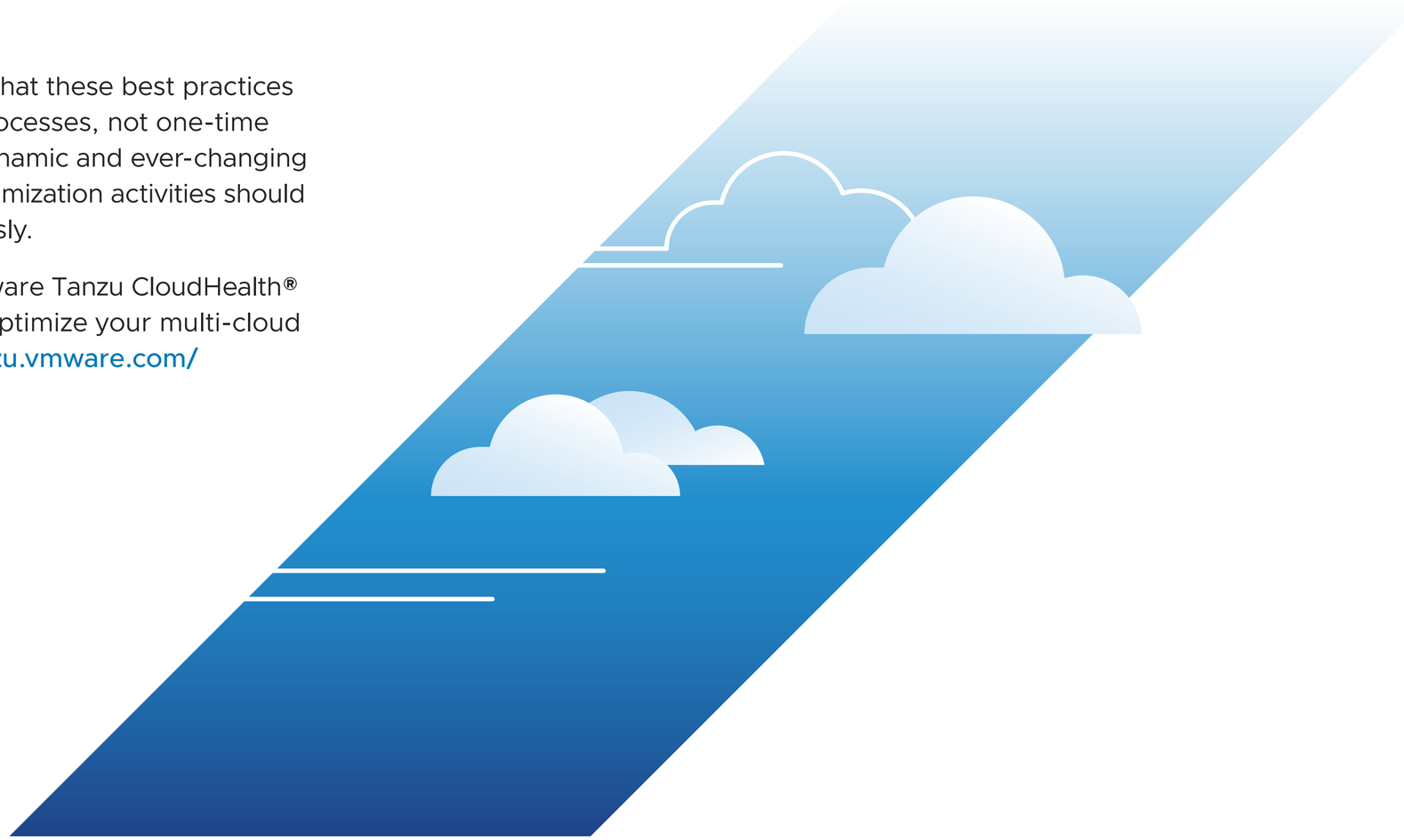
**Pro tip**

Many companies need higher storage classes for their new workloads. However, they forget to migrate to lower tiers of storage as the workload requirements subside. It is very important to keep track of this because the price differential from the top tier to the bottom is quite high. We recommend tracking the storage tier for your workloads on a frequent basis to save more.

# Conclusion

It's important to remember that these best practices are meant to be ongoing processes, not one-time activities. Because of the dynamic and ever-changing nature of the cloud, cost optimization activities should ideally take place continuously.

Learn more about how VMware Tanzu CloudHealth® can help you continuously optimize your multi-cloud environment by visiting tanzu.vmware.com/cloudhealth.



Get Started Today

## Try Tanzu CloudHealth for free.

LEARN MORE

Join us online: