Technical White Paper: 2022

Cloud Operations for Users of VMware Tanzu CloudHealth

Table of contents

About this white paper

This technical white paper is an accompanying piece to the Building a Successful Cloud Operations and Governance Practice white paper. In that white paper, we discuss how cloud operations is part of the cloud center of excellence (CCoE). By creating a CCoE, your company can start to successfully mature through the distinct phases of a cloud maturity journey: visibility, optimization, governance and automation, and business integration. In this technical white paper, we will share examples of how you can implement cloud operations best practices using Tanzu CloudHealth®.

Note: The examples included in this technical white paper are not an inclusive list and are merely a collection of commonly used reports and functionality among our customers.

What is cloud operations?

Cloud operations is the process of managing and delivering cloud services that meet the availability, performance, recoverability, quality and scalability needs of the business. Also referred to as CloudOps, this area of excellence manages the delivery, tuning, optimization and performance of workloads and IT services that run in a cloud environment, including multi-cloud, hybrid, in the data center, and at the edge.

CloudOps uses DevOps principles and IT operations as a cloud-based solution to improve business outcomes. The cloud operations team within an organization focuses on cloud optimization through key performance indicators (KPIs), rightsizing infrastructure, and creating structure for configuration and usage of resources. Establishing a fully functioning operations team requires ensuring operations meet and exceed business requirements, identify and act on areas to improve operational efficiency, and drive operational consistency across groups.

The goal of cloud operations, within the shared responsibility and security model of public cloud providers, is to ensure optimal functionality and performance of cloud platforms and associated workloads for their full lifecycle. The key term is optimal. In the centralized IT model, the metrics for optimal focus exclusively on reliability, availability and scalability (RAS). Centralized IT maintains and optimizes RAS by keeping tight control over resources. With public clouds, that approach is simply not possible and a fast track to failure.

Cloud operations best practices with Tanzu CloudHealth

Building the foundation for CloudOps

The ease of cloud services and the ubiquity of agile development processes and continuous integration/continuous delivery (CI/CD) frameworks are now the foundation of an enterprise's business and digital transformation. Moving to the cloud is one thing. Operating efficiently over time and maintaining alignment with business goals relies on embedding operational best practices, such as tagging hygiene, across the entire organization—and that is a much larger and more complex undertaking.

As organizations migrate workloads from the data center to the cloud, the need for CloudOps continues to grow. CloudOps encompasses cloud platform engineering principles, combining elements of cloud architecture, IT operations, application development, security, and regulatory compliance to enable organizations to manage cloud-based applications and services. It's recommended for DevOps and CloudOps teams to share best practices as they both promote improving customer user experience, enhancing productivity of teams, lowering cost of cloud delivery services, and growing the agile work environment for cloud workloads. DevOps improvements can bubble up throughout the organization, helping to bring more reliable software applications to fruition faster, which leads to improved performance for the organization. DevOps helps improve the user experience for employees and customers alike.

Phase 1: Visibility

The first step in assessing the state of your cloud environment is to get access to all accounts across all clouds. This initial step is a requirement for several other functions within the CCoE that also need visibility into cloud accounts to help the business better manage costs and ensure operational efficiency across environments.

If a business uses two or more clouds, then it often needs to use two or more cloud provider native tools, which may result in blind spots. Blind spots can prevent the identification of several issues, including unapproved shadow IT, unused or underutilized resources, inefficient application performance, unchecked security threats, and noncompliance with industry regulations. Visibility in the cloud means eliminating the blind spots that can lead to overspending, performance inefficiencies, and security issues. In an on-premises infrastructure, obtaining total visibility is easy; when your business operates in the cloud, it's that much harder. A cloud management platform, such as Tanzu CloudHealth, is an ideal solution to overcome the challenges of multi-cloud visibility by collecting data from all services used by the business. This is the only way to ensure you have the holistic view required to make well-informed decisions.

Resource tagging

A critical component of visibility stems from an organization's tagging strategy. Tagging strategies should be global to ensure resources are tagged consistently across all clouds. Once every resource is tagged, businesses have a starting point from which they can easily track, segment and analyze their cloud activities. Tag categories can be grouped based on cost, usage and security, to name a few. For example, common business tags include cost center, line of business (LOB), and business unit, whereas usage tags can include environment, function or project. Security tags can include compliance, public or confidential. An example of an advanced use case for tag management is that they can be leveraged to allow/deny access to certain resources. With proper tagging hygiene in place, businesses have a solid foundation as they focus on optimization and efficiency exercises, such as rightsizing.

Tanzu CloudHealth not only collects and collates data, its **Perspectives** capability enables customers to create unique business groups to view, analyze and evaluate their cloud environment from different lenses or points of view. Perspectives are defined by tag key-value pairs, naming conventions, accounts, services, and/or any metadata within the cloud environment. Each business group (e.g., finance, engineering, operations) can define its own Perspective to align cloud activity with the group's business metrics. Each Perspective contains groups and assets that can only belong to a group within a Perspective; however, the same assets can also be in a different Perspective (see Figure 1). Perspectives are available throughout the platform as filters in reporting, policies and recommendations.

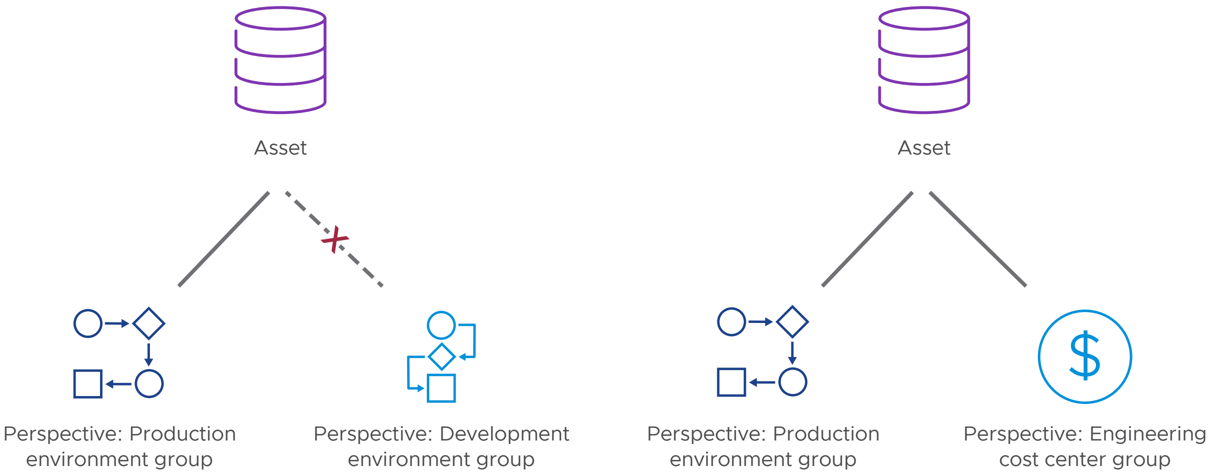

Asset
Asset
Perspective: Production environment group
Perspective: Development environment group
Perspective: Production environment group
Perspective: Engineering cost center group

Figure 1: An asset can only belong to one group within a Perspective, but it can also belong to a group within a different Perspective.

Tanzu CloudHealth tagging best practices

We recommend the following 10 multi-cloud tagging best practices to customers who are working to build out an organized and thoughtful approach to tagging:

1. Choose a standardized case-sensitive format as this helps to maintain the tagging architecture across different cloud providers.
2. Have a maximum of 50 tags per resource.
3. Have a maximum of 63 Unicode characters each for the key and the value.
4. Have only one value per each tag.
5. Plan for known restrictions for every cloud you are using or may use in the future.
6. Use tags to support the ability to easily manage resource access control, cost tracking, automation, and organization.
7. Apply more tags now in case they may be needed later. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) support at least 50 tags per resource.
8. Aim to implement a consistent standard across all cloud providers.
9. Enforce a standardization, such as ensuring all values are in camel case (e.g., CamelCase). This may prove beneficial, even though the tag or label key case sensitivity may differ from cloud to cloud. AWS, Azure, and GCP all have case-sensitive values.
10. Leverage Tanzu CloudHealth to tag resources that may otherwise not be supported as a taggable resource by the cloud provider. In addition, you can leverage both Tanzu CloudHealth tags and cloud provider tags when building Perspectives, which gives you greater flexibility in organizing your cloud infrastructure.

It's important that your business establishes tagging best practices and parameters. Read our blog post for an example.

FlexOrgs

With hundreds of cloud consumers across the distributed enterprise, consuming thousands of different services across multiple clouds, the scale of cloud operations is beyond the complexity that traditional tools or processes can handle. Tanzu CloudHealth FlexOrgs help solve this issue through increased delegation of administrative controls and responsibilities to users at various levels within the organization. You can provide the exact level of permissions to administrators, power users, and regular users at each level of the organization. Users can then create their own content, such as reports and policies that remain within their organizational unit (OU).

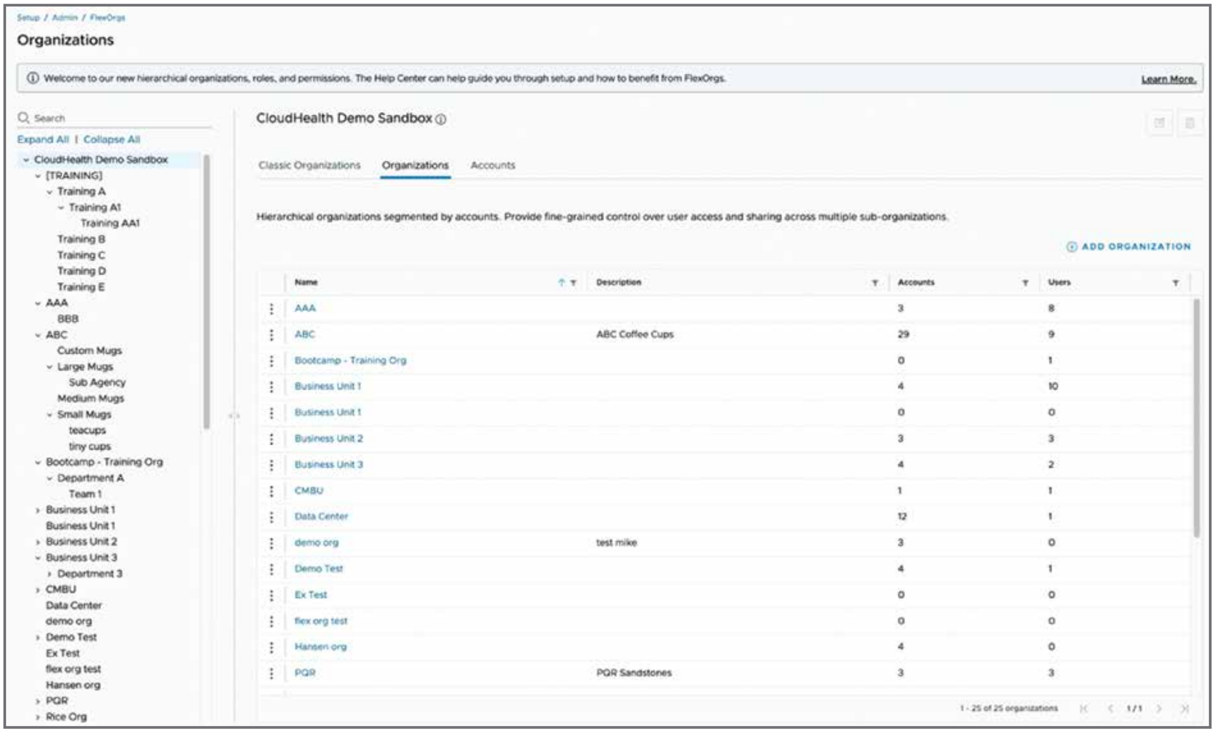
FlexOrgs provide greater control over user access, sharing and delegation across multiple levels of organizational hierarchy. FlexOrgs are made up of OUs, users, user groups, and role documents. OUs define what content and data an assigned user can see in the platform, and are based on accounts, subscriptions or projects depending on which cloud provider(s) you're using. Users are people who have access to the Tanzu CloudHealth platform. User groups define characteristics of an OU and assign permissions to users within the user group using role documents. User groups can give a user access to one or more OUs. Role documents are a list of permissions that define what a user can do with content in the platform. Role documents are attached to user groups.

With FlexOrgs, you build and manage your organization as a hierarchy composed of multiple OUs. Each OU is assigned access and responsibility for managing resources in the Tanzu CloudHealth platform.

FlexOrgs help you build organizations in Tanzu CloudHealth that closely reflect your business. You can assign a user to one, many or all FlexOrgs in the hierarchy, giving you more power over a user's permissions. For example, a user might have edit access to their primary organization but view-only access for their other organizations. It's important to know what your organization's end goal is when building FlexOrgs, so you know where to focus your efforts as you begin the building process.

Benefits of FlexOrgs

- Hierarchical representation – Define OUs that are responsible for managing specific portions of your cloud infrastructure. Link these OUs into a hierarchy that reflects how management control should be delegated.
- Segmentation – Segment your infrastructure by OUs in the hierarchy so that each unit has an isolated view of its portion of the infrastructure.
- User access – Map users to specific OUs and control their levels of access in each OU.



Setup / Admin / FlexOrgs

Organizations

ⓘ Welcome to our new hierarchical organizations, roles, and permissions. The Help Center can help guide you through setup and how to benefit from FlexOrgs. Learn More.

Search

Expand All | Collapse All

- CloudHealth Demo Sandbox
  - [TRAINING]
    - Training A
      - Training A1
        - Training AA1
    - Training B
    - Training C
    - Training D
    - Training E
  - AAA
    - BBB
  - ABC
    - Custom Mugs
    - Large Mugs
      - Sub Agency
    - Medium Mugs
    - Small Mugs
      - teacups
      - tiny cups
  - Bootcamp - Training Org
    - Department A
      - Team 1
  - Business Unit 1
  - Business Unit 1
  - Business Unit 2
  - Business Unit 3
    - Department 3
  - CMBU
  - Data Center
  - demo org
  - Demo Test
  - Ex Test
  - flex org test
  - Hansen org
  - PQR
  - Rice Org

CloudHealth Demo Sandbox ⓘ

Hierarchical organizations segmented by accounts. Provide fine-grained control over user access and sharing across multiple sub-organizations.

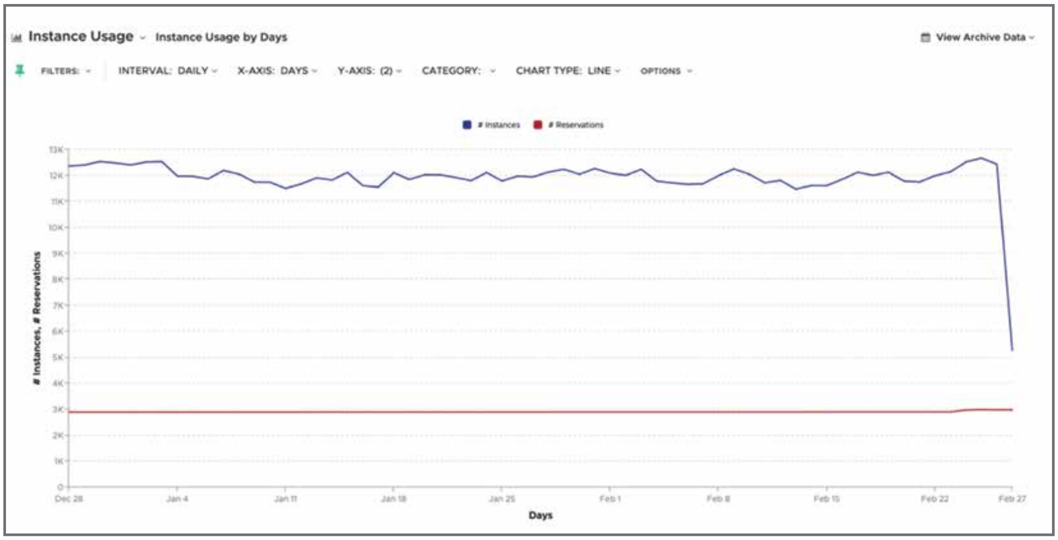
Name	Description	Accounts	Users
AAA		3	8
ABC	ABC Coffee Cups	29	9
Bootcamp - Training Org		0	1
Business Unit 1		4	10
Business Unit 1		0	0
Business Unit 2		3	3
Business Unit 3		4	2
CMBU		1	1
Data Center		12	1
demo org	test mike	3	0
Demo Test		4	1
Ex Test		0	0
flex org test		0	0
Hansen org		4	0
PQR	PQR Sandstones	3	3

1 - 25 of 25 organizations 1/1

Figure 2: An example of how organizations are created within the FlexOrgs capability.

EC2 Instance Usage Report

Once your tagging strategy is in place and Perspectives have been created, you can use Tanzu CloudHealth reports and dashboards to analyze your data and identify trends. As an example, let's look at the Amazon Elastic Compute Cloud (EC2) Instance Usage Report. This report looks at the number of instances running as compared to the overall cost of the instances running and is configured to categorize the data by reservation type. In Figure 3, you can see your total number of reservations compared to the number of instances running by selecting **# instances** and **# reservations** from the y-axis dropdown. You can then see how many instances and reservations you had over the past two months and check if the usage is consistent.



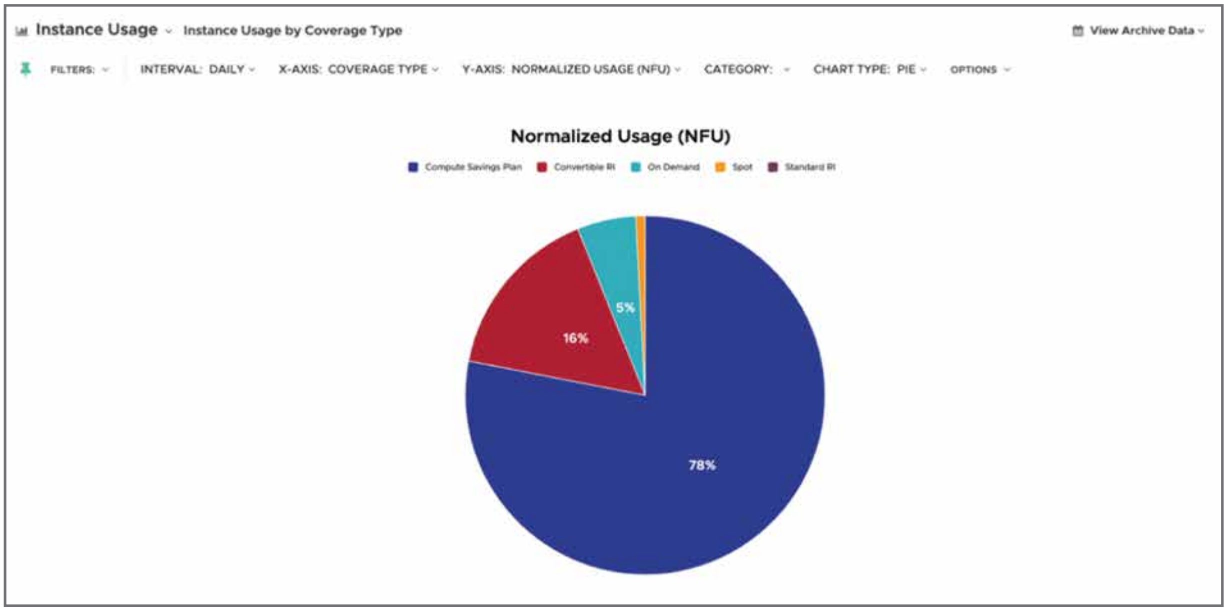
Instance Usage Instance Usage by Days

FILTERS: INTERVAL: DAILY X-AXIS: DAYS Y-AXIS: (2) CATEGORY: CHART TYPE: LINE OPTIONS

Figure 3: The Amazon EC2 Instance Usage Report.

This report helps you compare the size of reservations to the size of the instances running to see if there is the optimal amount of coverage for your infrastructure. These factors can be normalized to find any discrepancies between instances running and reservations coverage by selecting normalized reservations (NFU) and normalized usage (NFU) in the y-axis. This is an excellent way to see how you can optimize your infrastructure.

Another way that Tanzu CloudHealth provides visibility into your EC2 instance utilization is to see how your savings are applied by Reserved Instances (RIs) and Savings Plan (SP). In Figure 4, you can see your various coverage type ratios by changing the x-axis to coverage type and the y-axis to normalized usage (NFU), and then selecting pie as the chart type.



Instance Usage Instance Usage by Coverage Type

FILTERS: INTERVAL: DAILY X-AXIS: COVERAGE TYPE Y-AXIS: NORMALIZED USAGE (NFU) CATEGORY: CHART TYPE: PIE OPTIONS

Figure 4: Amazon EC2 instance usage by coverage type.

Phase 2: Optimization

Once an organization has passed through the visibility phase, the next phase of cloud management maturity is optimization. Most businesses understand the importance of optimizing their cloud infrastructure for usage, performance and cost-efficiency.

Rightsizing

Rightsizing is the process of matching instance types and sizes to your workload performance and capacity requirements at the lowest possible cost. Rightsizing can be defined in three steps:

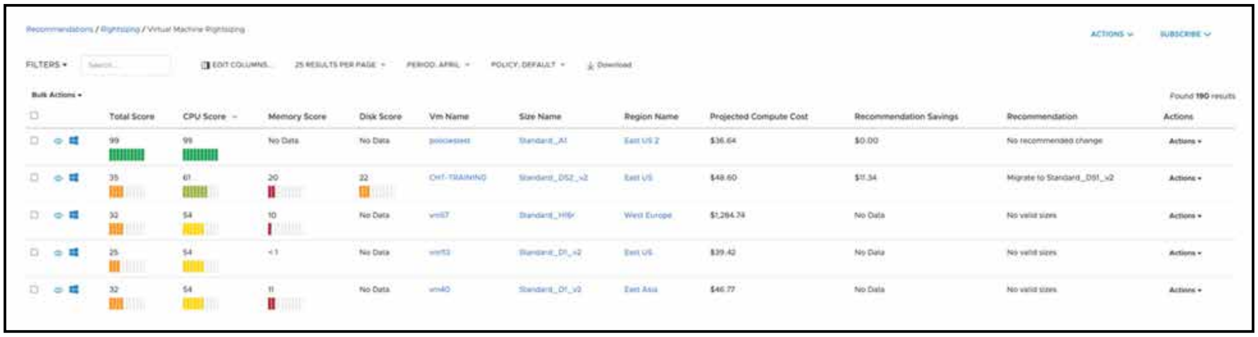
1. Analyzing the utilization and performance metrics of your infrastructure, such as instances, volumes and virtual machines (VMs)
2. Determining whether they are running efficiently, and what actions you should take to improve efficiency
3. Modifying the infrastructure as needed (upgrading, downgrading or terminating)

By rightsizing your cloud resources, you can benefit from infrastructure optimization and reduced costs.

The Tanzu CloudHealth platform supports rightsizing across Amazon EC2 instances, Amazon Elastic Block Store (EBS), Amazon Relational Database Service (RDS), Azure Virtual Machines, Azure SQL Databases, Google Compute Engine instances, and data center machines. The rightsizing recommendations from Tanzu CloudHealth are based on utilization and performance metrics (e.g., CPU, memory, disk, network) that can be ingested into the platform via an API, integration partners (e.g., Datadog), or the Tanzu CloudHealth Agent. Once the metrics are available, you have the power to set performance analysis thresholds specific to your applications and can take advantage of advanced filtering capabilities by dynamic business groupings, regions and more.

During a rightsizing analysis of their infrastructure, organizations usually discover assets that can be downsized or terminated to save money, or upgraded to improve performance and meet surges in demand. For example, in Azure, if you have a Standard_A2 VM with usage spikes that consistently hit 100 percent utilization of CPU or memory during certain times of the day, you may want to analyze the hourly maximum utilization throughout the day and see if the VM requires a larger size, such as the Standard_A3 or the burstable B-Series, to optimize performance. Organizations at the optimization phase are usually trying to utilize as close to 100 percent of the resources provisioned as possible while maintaining availability and performance across their workloads.

As an example, the Tanzu CloudHealth Virtual Machine Rightsizing Report provides various metrics, such as CPU, memory and virtual disk I/O, that contribute to VM utilization. It also shows how well your VMs are being utilized in terms of the workload you are running on them. Tanzu CloudHealth gathers metrics for each VM in your infrastructure. Each metric is then assigned a score, and the individual metric scores are used to compute a total score for each VM. While the individual metric scores indicate the utilization of each metric, the total score represents how well each VM is being utilized.



Recommendations / Rightsizing / Virtual Machine Rightsizing

Found 190 results

Total Score	CPU Score	Memory Score	Disk Score	Vm Name	Size Name	Region Name	Projected Compute Cost	Recommendation Savings	Recommendation	Actions
99	99	No Data	No Data	[illegible]	Standard_A1	East US 2	$36.64	$0.00	No recommended change	Actions
35	61	20	22	CHT-TRAINING	Standard_DS2_v2	East US	$48.60	$11.34	Migrate to Standard_DS1_v2	Actions
32	54	10	No Data	vm57	Standard_H16r	West Europe	$1,284.74	No Data	No valid sizes	Actions
25	54	<1	No Data	vm53	Standard_D1_v2	East US	$39.42	No Data	No valid sizes	Actions
32	54	11	No Data	vm40	Standard_D1_v2	East Asia	$46.77	No Data	No valid sizes	Actions

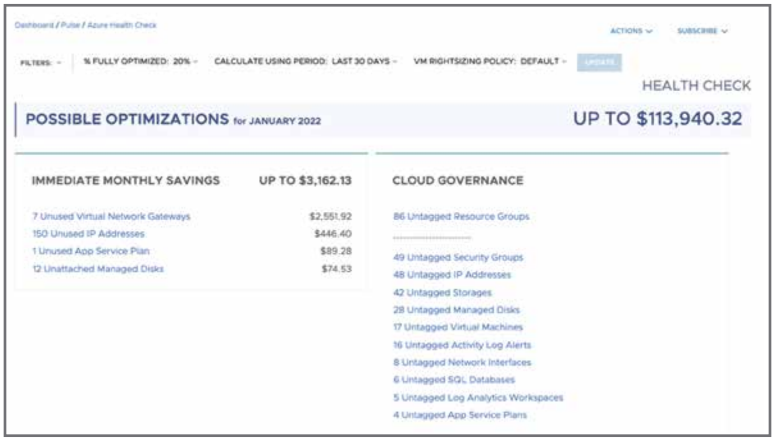
Figure 5: VM rightsizing recommendations.

Scores for individual metrics and the total score are visually represented as battery meters. Lower scores are indicated by fewer bars in red through orange. Larger scores are indicated by more bars in orange through green. Based on the recommendations and savings provided, you can make the decision to rightsize the resource.

Terminate zombie infrastructure

Another optimization best practice is to terminate assets that are considered zombies, which are assets running in your cloud environment but not being used. Zombies occur when someone may have forgotten to turn the assets off, or the asset failed because of script errors. Regardless of the cause, your cloud provider will continue charging for these assets because they are in a running state. By finding these assets and terminating them, you can reduce wasted cloud costs. Zombie assets come in many forms, such as compute infrastructure, databases, unattached storage volumes, disassociated IPs, and more. Within the Tanzu CloudHealth platform, zombies can be identified in several ways, such as by using the Health Check Pulse Report, rightsizing recommendations, or through governance policies.

The Tanzu CloudHealth Health Check Report is an executive assessment of your cloud infrastructure that highlights cost savings and cloud governance points. Unlike other Pulse reports, it is intended as a high-level report spanning several key capabilities. The immediate monthly savings section identifies different types of unused assets. For example, the report shown in Figure 6 identified virtual network gateways not being used by any resources and that can be deleted. Also, when you delete a VM, the IP address that was attached to it may persist unattached, generating unnecessary costs. These unused IP addresses can be deleted. The dollar amount to the right of the total number of IP addresses represents the total list price of unused IP addresses.



Dashboard / Pulse / Azure Health Check

% FULLY OPTIMIZED: 20%
CALCULATE USING PERIOD: LAST 30 DAYS
VM RIGHTSIZING POLICY: DEFAULT

HEALTH CHECK

POSSIBLE OPTIMIZATIONS for JANUARY 2022 — UP TO $113,940.32

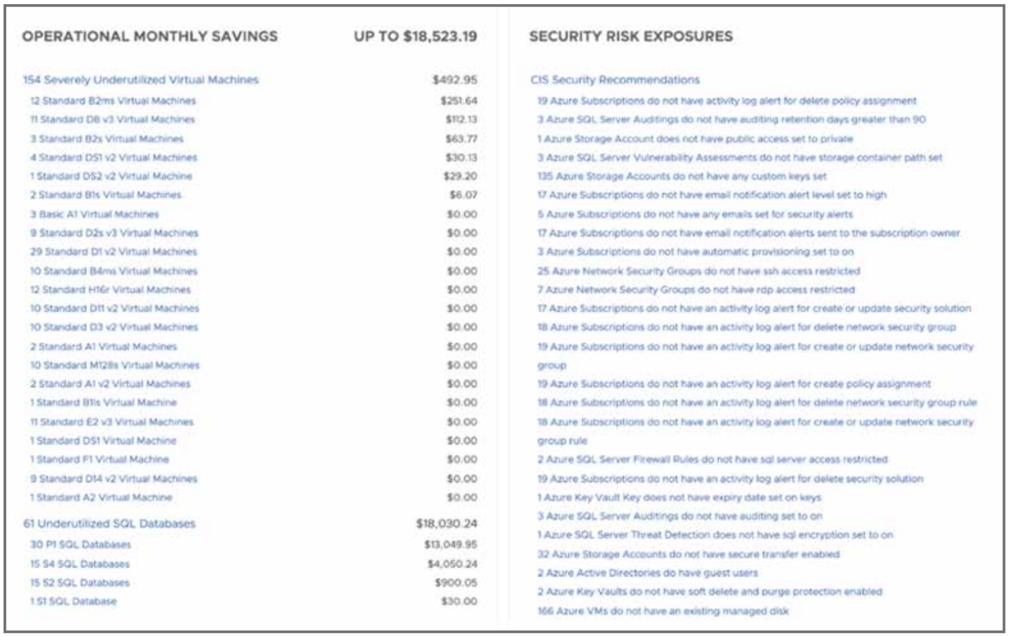
IMMEDIATE MONTHLY SAVINGS	UP TO $3,162.13
7 Unused Virtual Network Gateways	$2,551.92
150 Unused IP Addresses	$446.40
1 Unused App Service Plan	$89.28
12 Unattached Managed Disks	$74.53

CLOUD GOVERNANCE

86 Untagged Resource Groups
49 Untagged Security Groups
48 Untagged IP Addresses
42 Untagged Storages
28 Untagged Managed Disks
17 Untagged Virtual Machines
16 Untagged Activity Log Alerts
8 Untagged Network Interfaces
6 Untagged SQL Databases
5 Untagged Log Analytics Workspaces
4 Untagged App Service Plans

Figure 6: Azure Health Check Pulse Report with immediate monthly savings and cloud governance.

The cloud governance section identifies VMs and resource groups that have not been tagged and are more than likely not sanctioned by any business group. Common actions for untagged VMs are to either fix the tags or terminate them. The total savings represent the total cost, month to date, of the matching instances. Referring to the visibility phase, it's important to maintain your tagging hygiene over time.



OPERATIONAL MONTHLY SAVINGS	UP TO $18,523.19
154 Severely Underutilized Virtual Machines	$492.95
12 Standard B2ms Virtual Machines	$251.64
11 Standard D8 v3 Virtual Machines	$112.13
3 Standard B2s Virtual Machines	$63.77
4 Standard DS1 v2 Virtual Machines	$30.13
1 Standard DS2 v2 Virtual Machine	$29.20
2 Standard B1s Virtual Machines	$6.07
3 Basic A1 Virtual Machines	$0.00
9 Standard D2s v3 Virtual Machines	$0.00
29 Standard D1 v2 Virtual Machines	$0.00
10 Standard B4ms Virtual Machines	$0.00
12 Standard H16r Virtual Machines	$0.00
10 Standard D11 v2 Virtual Machines	$0.00
10 Standard D3 v2 Virtual Machines	$0.00
2 Standard A1 Virtual Machines	$0.00
10 Standard M128s Virtual Machines	$0.00
2 Standard A1 v2 Virtual Machines	$0.00
1 Standard B1ls Virtual Machine	$0.00
11 Standard E2 v3 Virtual Machines	$0.00
1 Standard DS1 Virtual Machine	$0.00
1 Standard F1 Virtual Machine	$0.00
9 Standard D14 v2 Virtual Machines	$0.00
1 Standard A2 Virtual Machine	$0.00
61 Underutilized SQL Databases	$18,030.24
30 P1 SQL Databases	$13,049.95
15 S4 SQL Databases	$4,050.24
15 S2 SQL Databases	$900.05
1 S1 SQL Database	$30.00

SECURITY RISK EXPOSURES

CIS Security Recommendations

19 Azure Subscriptions do not have activity log alert for delete policy assignment

3 Azure SQL Server Auditings do not have auditing retention days greater than 90

1 Azure Storage Account does not have public access set to private

3 Azure SQL Server Vulnerability Assessments do not have storage container path set

135 Azure Storage Accounts do not have any custom keys set

17 Azure Subscriptions do not have email notification alert level set to high

5 Azure Subscriptions do not have any emails set for security alerts

17 Azure Subscriptions do not have email notification alerts sent to the subscription owner

3 Azure Subscriptions do not have automatic provisioning set to on

25 Azure Network Security Groups do not have ssh access restricted

7 Azure Network Security Groups do not have rdp access restricted

17 Azure Subscriptions do not have an activity log alert for create or update security solution

18 Azure Subscriptions do not have an activity log alert for delete network security group

19 Azure Subscriptions do not have an activity log alert for create or update network security group

19 Azure Subscriptions do not have an activity log alert for create policy assignment

18 Azure Subscriptions do not have an activity log alert for delete network security group rule

18 Azure Subscriptions do not have an activity log alert for create or update network security group rule

2 Azure SQL Server Firewall Rules do not have sql server access restricted

19 Azure Subscriptions do not have an activity log alert for delete security solution

1 Azure Key Vault Key does not have expiry date set on keys

3 Azure SQL Server Auditings do not have auditing set to on

1 Azure SQL Server Threat Detection does not have sql encryption set to on

32 Azure Storage Accounts do not have secure transfer enabled

2 Azure Active Directories do have guest users

2 Azure Key Vaults do not have soft delete and purge protection enabled

166 Azure VMs do not have an existing managed disk

Figure 7: Azure Health Check Pulse Report showing operational monthly savings and security risk exposures.

As noted in the Rightsizing section, Tanzu CloudHealth provides rightsizing recommendations for several services across clouds. These recommendations are surfaced in the Health Check Pulse Report under the operational monthly savings section. This section identifies underutilized VMs and SQL databases to help you improve resource utilization and operational efficiency.

The security risk exposures section is based on the Center for Internet Security (CIS) Benchmark for Azure. This section identifies high and critical severity violations as defined in the default security policy.

Container reporting

Modern applications are increasingly being built using cloud containers because of their incredibly fast deployment speed, workload portability, and the ability to simplify resource provisioning for time-pressed developers. The explosion of container adoption over the past few years brought with it a wave of container-related products and services from the industry's leading cloud providers. For example, AWS has two fully managed container services: Amazon Elastic Kubernetes Service (EKS) and Amazon Elastic Container Service (ECS). Amazon EKS is a managed service that makes it easy to run Kubernetes on AWS without needing to install, operate and maintain your own Kubernetes control plane. Amazon EKS users are also able to take full advantage of the performance, scale, reliability and availability of the AWS platform. Further, Amazon EKS seamlessly integrates with a suite of other AWS services, and any application running on Amazon EKS is compatible with those applications already running in your existing Kubernetes environment. Amazon ECS is a fast, highly scalable, and fully managed container orchestration service that allows users to easily run and scale containerized applications on AWS. Unlike Amazon EKS, which is an AWS managed service for Kubernetes, Amazon ECS is AWS' own orchestration service that supports Docker containers.

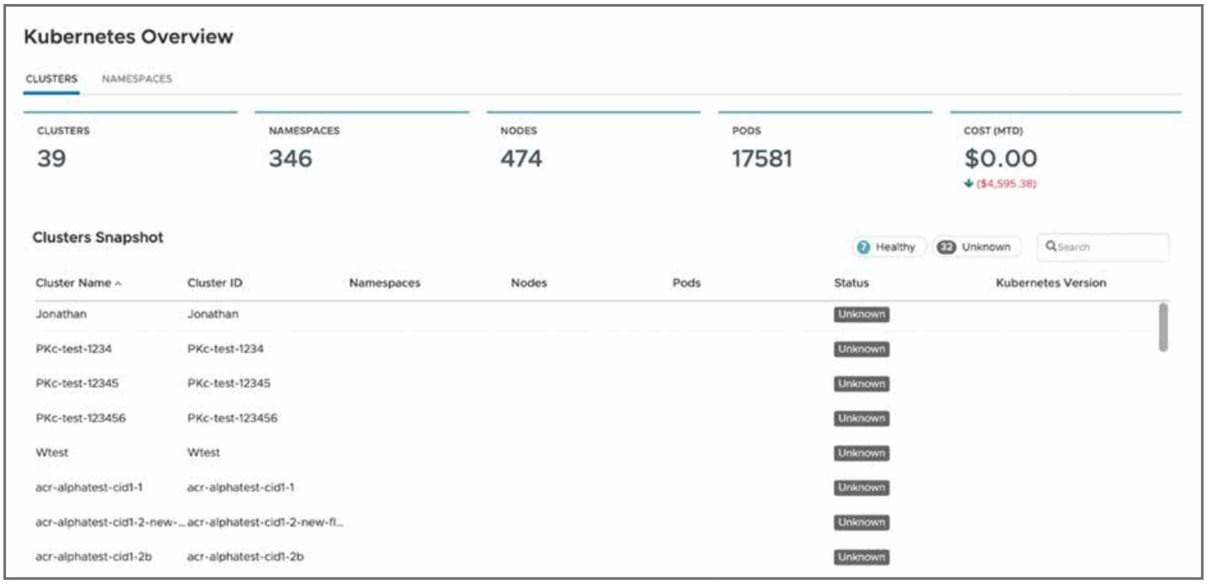
Kubernetes, the industry's leading container orchestration tool, makes it possible to deploy modern applications that are scalable, modular and fault tolerant. You declare the state you need your environment to be in and it constantly works to maintain that state, which frees developers from manual tasks around infrastructure management. With all the benefits of Kubernetes, there are also challenges when it comes to managing cloud costs. When more teams start to adopt containers and Kubernetes to develop and deliver their applications, your landscape can quickly become crowded and fragmented. This is where a CCoE comes in. The CCoE can bring together key stakeholders across your organization—from development, finance and operations—for a unified approach to cloud management. A CCoE can help identify the most important KPIs to the business, and then align different business units to ensure everyone tracks the same metrics and utilizes cloud resources for the same bottom line.

Once your organization has aligned on a KPI, it's important to standardize definitions and labels in your Kubernetes cloud infrastructure. As a best practice, you can set up automated governance to ensure consistency. Once you have defined labels and definitions, you can start to think about how to group and allocate costs, which can be done either by the provisioned amount or the used amount. Which model you choose depends on the organization. In either case, the principle holds that if your usage is lower than what you request, you're wasting money. If it's higher, you risk development or performance issues. That's why it's important to have the right tools in place and visibility into your environment.

Containers are revolutionizing how applications are developed and deployed, and can provide numerous advantages for organizations looking to adopt the technology. However, optimizing your Kubernetes cloud costs can be difficult without visibility into your containerized environment and accountability for spend. To help solve this challenge, Tanzu CloudHealth provides container reporting capabilities that give customers visibility into their Kubernetes and container environments, and help identify areas for optimization.

For example, the Kubernetes Overview Report consolidates key Kubernetes reports in the platform to one summary-level report. You can choose to view costs and usage either by cluster or namespace, as shown in Figure 8. You can see at the top of the report a count of clusters, namespaces, nodes and pods that the Tanzu CloudHealth platform collects.

The cluster snapshot provides information regarding the configuration of your clusters. The status of these clusters is shown in a healthy or unhealthy state. Unknown status means the collector has not successfully contacted Tanzu CloudHealth.



Kubernetes Overview

CLUSTERS	NAMESPACES	NODES	PODS	COST (MTD)
39	346	474	17581	$0.00 ($4,595.38)

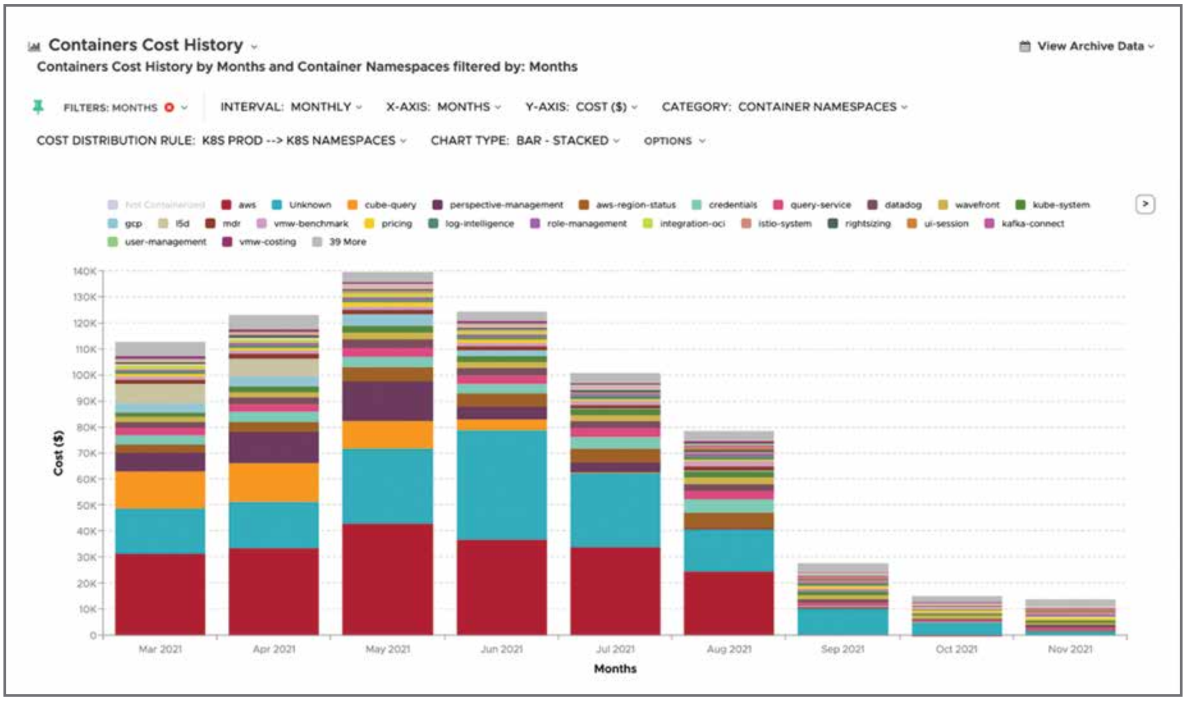
Clusters Snapshot

Cluster Name ^	Cluster ID	Namespaces	Nodes	Pods	Status	Kubernetes Version
Jonathan	Jonathan				Unknown	
PKc-test-1234	PKc-test-1234				Unknown	
PKc-test-12345	PKc-test-12345				Unknown	
PKc-test-123456	PKc-test-123456				Unknown	
Wtest	Wtest				Unknown	
acr-alphatest-cid1-1	acr-alphatest-cid1-1				Unknown	
acr-alphatest-cid1-2-new-...	acr-alphatest-cid1-2-new-fl...				Unknown	
acr-alphatest-cid1-2b	acr-alphatest-cid1-2b				Unknown	

Figure 8: Kubernetes overview.

Another example is the Containers Cost History Report, which can be used to visualize cost trends in aggregate, and per cluster and namespace. You can configure multiple concurrent distributions of costs used by containerized applications and measure the costs of a cluster; for example, split by namespace, product line or team. This capability allows users to select their preferred cost distribution rule in the Containers Cost History Report from a dropdown selector.

In Figure 9, the Container Namespaces Perspective has been selected from the category dropdown. Users can select their preferred cost distribution rule from the dropdown selector. In this example, this rule is taking the cost of the resources supporting the Prod cluster and distributing those costs to the namespaces.



Containers Cost History

Containers Cost History by Months and Container Namespaces filtered by: Months

FILTERS: MONTHS

INTERVAL: MONTHLY

X-AXIS: MONTHS

Y-AXIS: COST ($)

CATEGORY: CONTAINER NAMESPACES

COST DISTRIBUTION RULE: K8S PROD --> K8S NAMESPACES

CHART TYPE: BAR - STACKED

Figure 9: Containers Cost History Report.

Phase 3: Governance and automation

After you have identified and optimized areas for improvement, you can implement governance policies and eventually automate daily tasks. When a business operates in the cloud, the rules of operation can be much more difficult to enforce. Due to the self-provisioning nature of the cloud, costs are now fluid operating expenses, software and apps can be downloaded with the click of a mouse, and there are no firewalls to protect the business's data from theft or unauthorized disclosure. Costs can spiral out of control, LOB shadow IT environments develop, and there are no controls over where data is or how it is protected. This places the business at risk of data loss as well as at risk of noncompliance with industry privacy and security standards.

Your organization should establish standard operating procedures and rules. The rules should cover the what, when and how of operating in the cloud. For example, what cloud services can be used, when users should download other software and apps (usually involving an approval process), and how data must be protected. Everyone needs to understand why these cloud computing governance rules are in place. There are two main ways to enforce cloud computing governance rules: manually and automatically. The manual option involves constantly monitoring the cloud environment for breaches of the rules or events that could lead to a breach of the rules. Automatic cloud computing governance enforcement involves comparatively minimal effort once you understand automation and if your operating procedures allow for automation.

There are two types of governance policies:

- Guidelines – Guideline policies will communicate a risk boundary via an alert that informs the user of the best practice but will not take action to prevent or correct the action.
- Guardrails – Guardrail policies will both communicate and take action to correct a violated best practice (e.g., terminating zombie infrastructure).

In most use cases, automated governance solutions are configured to send email notifications, for example, in the event of a potential budget overspend, when there is an opportunity to take advantage of a commitment-based discount, or when performance metrics indicate an asset requires rightsizing. Good examples of guardrails for developers include policies that deny accidental changes to baseline security monitoring controls or those that require boundary permissions for all identity and access management (IAM) users and roles. While guardrails are nondisruptive, security teams must communicate these to developers before implementation. This helps ensure developers are not surprised by new policies, understand why they exist, and help security teams to implement them.

Tanzu CloudHealth can be used to implement both guideline and guardrail policies. For example, if a VM is assigned a non-U.S. region, exceeds the permitted maximum capacity, or is launched outside normal working hours, Tanzu CloudHealth automatically terminates the asset. Similarly, Tanzu CloudHealth can revoke access to accounts that launch assets out of the normal range, out of normal working hours, or that are logged in to from an unrecognized IP address. Tanzu CloudHealth collects data from all the services being used by the business to provide total visibility of the IT infrastructure. Once the data has been aggregated, you can use policy-driven automation to police the cloud governance framework and ensure your governance rules are being adhered to.

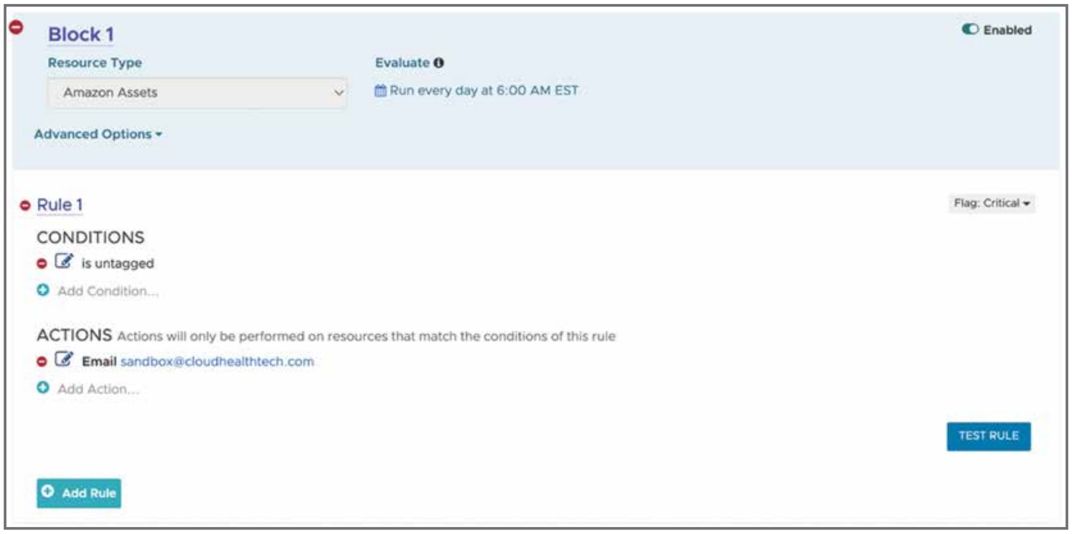
Tag compliance by asset

Cloud operations teams are often ruthless about tagging compliance. Tagging is an essential way to accurately group assets in their appropriate business groups. You can set notifications in Tanzu CloudHealth to identify assets that do not conform with the internal tagging standards defined by your organization. Not tagging assets or not following tagging rules makes it impossible to associate an asset with an owner, business unit, cost center, and the like. Put simply, poor tagging compliance nearly renders the asset unmanageable. This is one of the instances where an out-of-band guardrail is required to automatically notify owners of improperly tagged assets and inform them they will automatically be deleted after
a set time if not remediated.

Examples include:

- If any asset is missing the Environment tag, send a notification and execute a lambda function to tag the asset.
- If any asset is untagged, alert its owner and stop the instance.

The policy shown in Figure 10 sends an email alert reporting any new AWS assets that are provisioned without being tagged.



Block 1

Enabled

Resource Type: Amazon Assets

Evaluate: Run every day at 6:00 AM EST

Rule 1

Flag: Critical

CONDITIONS

is untagged

ACTIONS Actions will only be performed on resources that match the conditions of this rule

Email sandbox@cloudhealthtech.com

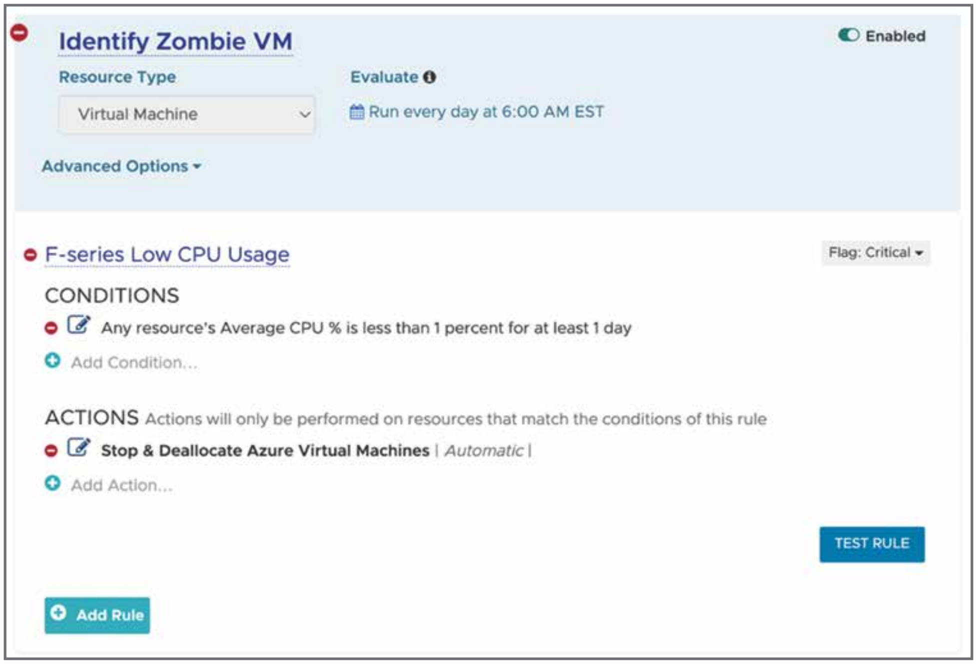
Figure 10: AWS asset governance policy.

Identify and terminate zombie VMs

Zombie VMs are running virtual machines that are idle, mostly forgotten, and cost you money. With Tanzu CloudHealth, you can identify VMs running with a daily average CPU rate lower than 10 percent for two weeks in a row and network I/O less than 5MB for four or more days. If you want to be more specific, isolate instances based on their instance type.

For example, F-series VMs (compute optimized) that have a maximum CPU rate of less than 10 percent for the past 14 days are most likely to be running idle and are good candidates to be terminated.

The policy shown in Figure 11 identifies F-series VMs (compute optimized) that have a low average CPU rate and sends a notification, as well as specifies a course of action you can take. In addition, by leveraging Perspectives, you can run this policy against specific non-production environments.



Identify Zombie VM

Enabled

Resource Type: Virtual Machine

Evaluate: Run every day at 6:00 AM EST

Advanced Options

F-series Low CPU Usage

Flag: Critical

CONDITIONS

Any resource's Average CPU % is less than 1 percent for at least 1 day

Add Condition...

ACTIONS Actions will only be performed on resources that match the conditions of this rule

Stop & Deallocate Azure Virtual Machines | *Automatic* |

Add Action...

TEST RULE

Add Rule

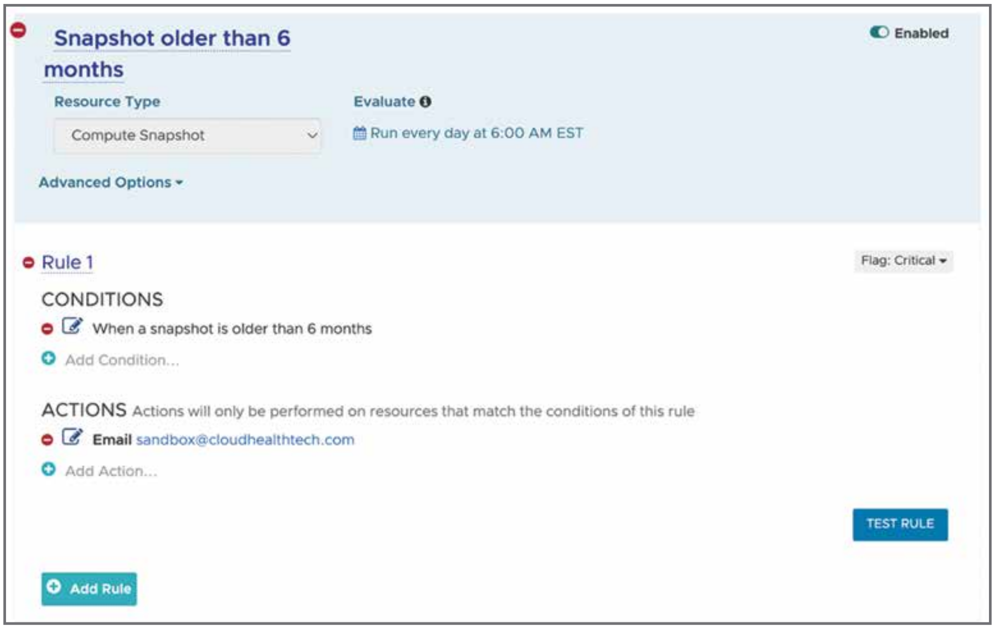
Figure 11: Azure zombie VM identifying policy.

Identify and delete old snapshots

Compute snapshots can be valuable storage tools but are often left forgotten and continue to incur cost as they age. Old compute snapshots can also become a legal liability when storing sensitive or regulatory data.

For example, you can identify snapshots that are older than six months and terminate them after confirming they do not contain critical data.

The policy shown in Figure 12 sends a notification when it identifies potential zombie compute snapshots that are older than six months.



Snapshot older than 6 months

Enabled

Resource Type: Compute Snapshot

Evaluate: Run every day at 6:00 AM EST

Rule 1

Flag: Critical ▾

CONDITIONS

When a snapshot is older than 6 months

ACTIONS Actions will only be performed on resources that match the conditions of this rule

Email sandbox@cloudhealthtech.com

Figure 12: GCP old snapshot identifying policy.

Phase 4: Business integration

By the final phase in your cloud maturity journey, your business should have complete visibility into all your cloud environments, segmented by various business groups, as well as be continuously monitoring, optimizing and governing cloud resources. The last step is to integrate your cloud and business processes. One way to do so is by establishing KPIs that can be easily tracked and aligned with the company. These KPIs should be cross-functional, which ensures that everyone is working toward the same goal. CloudOps teams should align their cloud and operational KPIs with business KPIs.

Aligning metrics

Cloud management metrics can be used to provide business context to IT and engineering teams, and allow business users to easily understand the impact of the cloud center of excellence. These metrics could be about cloud costs and how they contribute to cost of goods sold (COGS), or as cost per customer support. Know what metrics can be used as drivers for tracking strategic initiatives. These metrics should stay constant even as businesses scale.

Business KPIs

With all the optimization and governance efforts you are making across the different functional areas of your organization, it is important that you can clearly show how your cloud strategy is driving business-wide transformation and impacting your organization’s corporate goals.

Here are some examples of KPIs to ensure your cloud strategy and business strategy are aligned:

- Cost per customer
- Cloud spend as a percentage of revenue
- Reduction in COGS over time
- Cost of revenue over time
- Time to bring new services to market
- Mean time to detect
- Customer satisfaction (typically using a Net Promoter Score)

Integrations and APIs

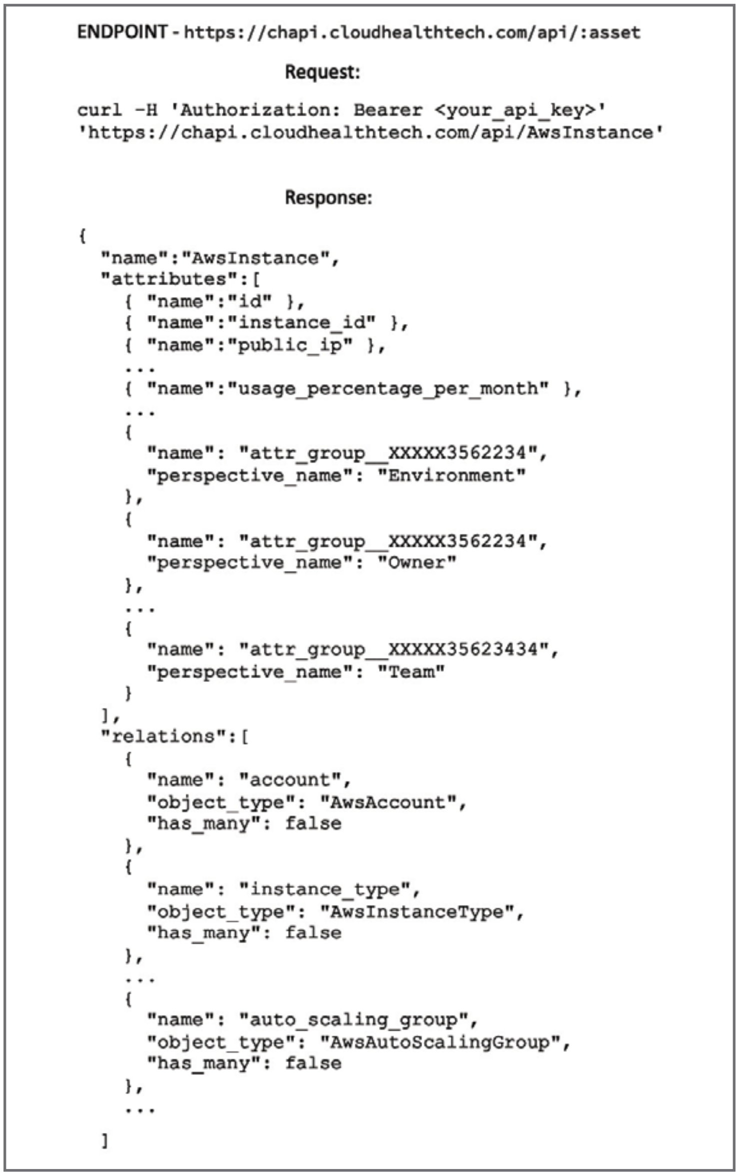
Tanzu CloudHealth is designed to leverage your existing toolset across performance management, configuration management, and security management to support your business needs. For example, Tanzu CloudHealth's integration with Datadog provides a better understanding of performance trends and requirements, so you can make good provisioning decisions. Once your account is set up, the imported Datadog tags can be used in Perspectives and reports.

With Tanzu CloudHealth APIs, customers can initiate requests to view, create or modify data within the Tanzu CloudHealth platform. These requests can come from any source, including a service request management system. Tanzu CloudHealth provides a representational state transfer (REST) API and a GraphQL API for programmatically interacting with the platform. The REST API has predictable, resource-oriented URLs, and it uses HTTP response codes to indicate API errors. All API responses, including errors, return JSON. The goal of the Tanzu CloudHealth API is to let you write your own applications that leverage and extend Tanzu CloudHealth functionality. You can send REST API requests to various endpoints to retrieve and update data from the Tanzu CloudHealth platform. Different API functionality can be found at different endpoints.

For example, the Assets API allows you to retrieve information on AWS, Azure, data center and Google Cloud assets in your environment. The API supports the following operations:

- Retrieve the API names of all AWS, Azure, data center and Google Cloud asset objects that you can query in the Tanzu CloudHealth platform.
- Retrieve the attributes of an asset, including the Perspective groups to which the asset belongs, as well as assets related to the queried asset.
- Query objects of a specific type, filter objects by attribute, and list specific fields of assets in the response.

Figure 13 shows an example that retrieves the attributes and related assets for a single asset object using GET request.



ENDPOINT - `https://chapi.cloudhealthtech.com/api/:asset`

Request:

```
curl -H 'Authorization: Bearer <your_api_key>'
'https://chapi.cloudhealthtech.com/api/AwsInstance'
```

Response:

```
{
  "name":"AwsInstance",
  "attributes":[
    { "name":"id" },
    { "name":"instance_id" },
    { "name":"public_ip" },
    ...
    { "name":"usage_percentage_per_month" },
    ...
    {
      "name": "attr_group__XXXXX3562234",
      "perspective_name": "Environment"
    },
    {
      "name": "attr_group__XXXXX3562234",
      "perspective_name": "Owner"
    },
    ...
    {
      "name": "attr_group__XXXXX35623434",
      "perspective_name": "Team"
    }
  ],
  "relations":[
    {
      "name": "account",
      "object_type": "AwsAccount",
      "has_many": false
    },
    {
      "name": "instance_type",
      "object_type": "AwsInstanceType",
      "has_many": false
    },
    ...
    {
      "name": "auto_scaling_group",
      "object_type": "AwsAutoScalingGroup",
      "has_many": false
    },
    ...
  ]
```

Figure 13: Asset API example.

The response to this query contains two arrays: attributes and relations. The attributes array lists the primary attributes of the cloud asset that you want to explore. For example, an AWS instance has attributes such as Account ID, Instance ID, Public IP and Private IP. The attributes array also includes the Perspective groups to which the AWS instance belongs. The relations array contains related assets. In the case of an AWS instance, the relations array lists objects such as AwsAccount and AwsInstanceType.

The Tanzu CloudHealth next-generation API is based on GraphQL. This API allows you to send a single query to the GraphQL server that includes specific requirements on which reporting fields you want included and how you want them organized. The server responds with a JSON object that fulfills those reporting requirements. This API only exposes a single endpoint because the structure of the reporting data that the GraphQL server returns is not fixed. In fact, there is a lot of flexibility in terms of defining what the structure of the reporting data should be.

Note: The documentation for the GraphQL API is currently not available for public viewing and requires a Tanzu CloudHealth login to access. The documentation for the REST API can be found at https://apidocs.cloudhealthtech.com/.

Conclusion

Every organization will define and structure a CCoE a bit differently; there is no one-size-fits-all approach. In any case, we recommend that cloud operations be one of the three areas of excellence for your CCoE, alongside cloud financial management and cloud security and compliance.

While cloud service providers own most of the traditional IT operations functions, it is a shared responsibility model in which the organization's cloud operations team ensures optimal functionality, cost and performance across all cloud platforms and associated workloads. Tanzu CloudHealth can help businesses drive operational consistency across the four phases of cloud management maturity.