Manage Kubernetes Costs with VMware Tanzu CloudHealth

Why use Tanzu CloudHealth?

- Understand which teams, departments or applications are driving Kubernetes costs.
- Perform showback or chargeback of shared cluster costs.
- Empower app developers to optimize Kubernetes requests and limits.
- Get recommendations to optimize Kubernetes resources and save money.

Kubernetes wins the container orchestrator race

The verdict is in, and Kubernetes has emerged as the de facto container orchestrator for businesses across the globe. According to the Cloud Native Computing Foundation (CNCF), 96 percent of organizations are either using or evaluating Kubernetes.¹ Further, there are more than 109 tools on the market to manage containers, and 89 percent of them use Kubernetes.² Kubernetes is taking the software development world by storm as it makes building and maintaining applications faster and easier.

Challenges with Kubernetes cost management

With new technology always comes new challenges. Containerization adds a layer of abstraction to your cloud environment that makes it hard to see the applications using resources and incurring costs. Your cloud bill will show a total charge per compute resource (e.g., node), but because each node powers several containers, you have no idea how much compute power each container used, or how much it costs.

Kubernetes also creates a capacity management challenge for application developers as Kubernetes asks them to reserve capacity from available nodes (i.e., set a request), and set a limit on the maximum amount of capacity the container can use at any given time. The challenge for developers is accurately estimating the amount of capacity the container will require.

How Tanzu CloudHealth can help

Tanzu CloudHealth® supports Kubernetes-based workloads across all environments, including public and private clouds and private data centers. It also supports Kubernetes managed services, such as Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), and Google Kubernetes Engine (GKE).

^{1.} Cloud Native Computing Foundation. "Annual Survey 2021."

^{2.} DevopsCube. "<u>16 Best Container Orchestration Tools and Services</u>." Bibin Wilson. January 5, 2022.

Manage Kubernetes Costs with Tanzu CloudHealth

"[Tanzu CloudHealth] lets us maximize the value of our container deployment by telling us how well utilized it is and whether our clusters have the right mix of resources supporting it. That level of insight enables us to make informed, strategic business decisions without additional overhead."

Director of Engineering Yelp The lightweight Tanzu CloudHealth container agent can be deployed to each cluster in your environment to gather information about available resources and how they are allocated to workloads. You can onboard Kubernetes clusters in bulk with a Helm chart or YAML file, and monitor the status of each agent to ensure usage metrics are being collected.

Get visibility into Kubernetes costs on any cloud

Tanzu CloudHealth maps shared Kubernetes resource costs back to your unique business groupings. You can see costs broken down by different groupings you define, such as applications, teams, departments and cost centers. The platform also supports breaking down costs by Kubernetes namespaces, namespace labels, and pod label groupings.

FinOps professionals in your organization can provide financial managers with Kubernetes costs per cost center or business unit for chargeback or showback, and for accounting, forecasting, and budget planning:

- Get visibility into Kubernetes resource utilization and costs across environments.
- Perform showback or chargeback of shared cluster costs by namespace, label, or unique business grouping.
- Onboard clusters in bulk with Helm chart and monitor data collection status.



Figure 1: Container costs broken down by business unit.

Manage Kubernetes Costs with Tanzu CloudHealth

Optimize Kubernetes resources to save money

Tanzu CloudHealth looks inside your Kubernetes clusters and determines how much each container uses a shared resource. FinOps teams can share this granular usage data with application developers, giving them the information they need to determine the appropriate CPU/memory requests and limits for their containers. If their container usage is consistently below the requested (or reserved) amount, they can lower the request amount. If all the containers in an environment request less capacity, then the business doesn't need as many nodes, so you can spend less money on compute instances:

- Track CPU and memory usage against requests, limits and available capacity.
- Highlight usage patterns ripe for optimization.
- Automate delivery of usage reports to inform DevOps teams.
- Empower application owners to make decisions about resource optimizations.



Figure 2: Empower application developers to optimize Kubernetes requests and limits.

Manage Kubernetes Costs with Tanzu CloudHealth

Act on rightsizing recommendations

Tanzu CloudHealth analyzes past usage data and generates recommendations to optimize the cost of your Kubernetes workloads by adjusting CPU and memory requests, or reserved capacity, on available nodes. Default recommendations are based on the history of your workload needs and industry benchmarks. Users can customize these recommendations by setting specific targets for CPU and memory:

- Receive recommendations for adjusting requests.
- Customize recommendations with custom efficiency targets.
- Filter by cluster, namespace or workload.
- Limit cluster access and permissions with FlexOrgs.
- Ensure optimal performance and reduce resource waste.

									ACTIONS
tightsizing									Manage efficiency tar
C2 RD5 Kubers	etes								Add custom efficiency
⊻ FILTERS Aver	ncy Target nge Mintrics 🛩 🗇	Date Range Last 7 Days . V	ATE REEF						
Summery									
Current CPU		Recommended CPU		Current Memory		Recommended Memory		nçe.	Resize
897.67	31	316.08-388.37		3221.80		3024.73-3629.68			2490
Requests (Coves)		gemta (Cores)	Recu	Requests (08)		Nequesta (5/8)		ria -	Containers
bernetes Recon	nmendations								
bernetes Recon	nmendations	November Serve	Generator Tree	CPUBments	Marrow Towards		Terost 🐻 (Over Target)	Ouderserviced Per	D Q Search nome, ID or to
Demates Recon	nmendations xPORT	Namespace Name	Becore Type DEPLOYNENT	CPU Requests	Memory Requirie	(Coost 711 () (Under CRU Ang Lange () 55 440%	Terost () (Over Target Becommended E 4 8-41-11 29 Cares	Underworked 7th Recommended Harns, 127.75-55.26.05	O O Search name, ID or to Economication Result CPU and memory requests
Den al ()	Cuter Name	Newsgard Name	Becarts Type DEPLOYMENT DEPLOYMENT	CPU Begants 10,00 Cures 0.50 Cures	New or Reports 119.00.08 2.00.08	(Coost TH ()) (Linder CPU Arg Usage (2) 56.46% (2) 786.2%	Terrant (Over Target) Becommended C. 4 8.41/11.29 Cures 8.55.7.86 Cores	Undeterwined 7tt Becommond Harrs. U7.71-153.26.08 2.92-3.52.08	P. Q. Search name, ID or to Recommodern Repart CPU and memory requests Bable CPU and memory requests.
Dernetes Recon	nmendations	Nesspectrate	Beseever Type DEPLOYMENT DEPLOYMENT	CPU Begarts 10.00 Cares 0.50 Cares 0.50 Cares	Marroy Zegato 19:00:08 2:00:08 2:00:08	(Count Thi () (Linder CRU Ang Unage () 56.465 () 784.275 () 444.875	Cover Target Cover Target Reconversed C. 4 S.4/it 23 Cover ESS-7.85 Cover S.T-64.85 Cover	Ordene worked 711 Becommenced Marris, 127,7h-153,26,08 2.99-3,52,G8 5.32-6.38,08	D Q Search name, ID or to BournmandAon Bourn CPU and mamory requests Bourn CPU and mamory requests Bourne CPU and mamory requests
Dennetes Recon	Current Rame	Newspect Name	Become Type DEPLOTMENT DEPLOTMENT DEPLOTMENT DEPLOTMENT	CPU Begants 10.00 Cares 0.50 Cares 0.50 Cares 0.50 Cares	Marroy Reside 119.00.08 2.00.08 2.00.08 2.00.08	(Coost TH () (Linde SRU Ang Lange 2: 55 46% () 796 2% () 464 87% () 55 368%	Cover Tanael Cover Tanael Reconverted E 4 8.4/1129 Cares 8.55 7 86 Cove 5.11-6.85 Cove 4.60-5.52 Cove	Cludenerwined 7t Recommended Hams. 127.7h:55.26.05 2.93.3.52.GB 5.32-6.38.05 3.00-3.00.05	D C, Search name, D or to Reservance
Dennetes Recon	Cuter Name	Newspect form	Benever Type DEPLOTMENT DEPLOTMENT DEPLOTMENT DEPLOTMENT DEPLOTMENT	CPU Breamb 10,00 Curve 0,50 Curve 0,50 Curve 0,50 Curve 4,00 Curve	Marroy Repairle 119.00.08 2.00.08 2.00.08 2.00.03 46.00.08	(Count Till () (Under CRU Ang Lange Cli 56.465 () 756.25 () 664.675 () 557.695 () 66-355	Cover Tanset Cover Tanset Cover Tanset Set#129 Cores S57 85 Cores S714.85 Cores 460-552 Cores 445-25 Cores	Cundeneword 715 Recommended Martine, 27 79-155 26 08 2 99-3 52 08 5 32-6 38 08 3 00-3 00 08 59-79-77 02 08	D O Search nome, ID on to Paramentation Result OV and memory requests Res
Ubernetes Recon	Cuter Karte		Beautre Type DEPLOYMENT DEPLOYMENT DEPLOYMENT DEPLOYMENT	CPU Begants 10,00 Cures 0,50 Cures 0,50 Cures 4,00 Cures 4,00 Cures	Menory Research 19:00:08 2:00:08 2:00:08 45:00:08 84:00:08	(Coost TH) (Unew CHU Ang Leage CHU Ang Leage Chu Shada Una Chu Shada Un	Cherr Tarraell	Childen wined 71 Becommodel Maria, 77-152 26 08 193-3 52 08 193-3 52 08 100-3 50 08 100-3 50 08 100-3 50 08 10-3 50 08 10-3 50 08 10-3 50 08	P Q Search some Drive to Research and memory requests Research and memory requests Research and memory requests Research and memory requests Research and memory requests

Figure 3: Customize recommendations for rightsizing Kubernetes requests.

As Kubernetes continues to gain market share and grow in cloud environments, businesses need a comprehensive platform that can manage Kubernetes workloads alongside traditional public cloud deployments. Tanzu CloudHealth provides end-to-end support for Kubernetes workloads in any environment, helping businesses deliver modern digital services while driving down costs.

Learn more

Ready to manage Kubernetes costs as part of a holistic cloud cost management strategy? Visit us online today and <u>try Tanzu CloudHealth for free</u>.