



AWS Cost Optimization

Best practices to lower your cloud bill

Table of contents

Executive summary	3
Challenges with cloud cost management	3
AWS Well-Architected Framework: A foundation for cloud success	4
Well-architected pillar no. 5: Cost optimization	4
In practice: Cost optimization for the cloud	5
1. Practice cloud financial management	5
2. Expenditure and usage awareness	7
3. Cost-effective resources	8
4. Manage demand and supply resources	11
5. Optimize over time	13
Conclusion	14

Executive summary

Enterprises rely on Amazon Web Services (AWS) to deliver agility, operational flexibility and faster innovation. However, managing cloud costs and operations is significantly different from managing on-premises data center costs. It can be difficult for IT organizations to achieve the visibility they need to not only optimize cloud costs and resources, but to align cloud spend with business initiatives.

To address these challenges, many organizations are adopting cloud financial management (CFM). The CFM approach is essential for organizations seeking to improve their cloud operations and reduce wasted spending, especially as those operations increase in scale.

The AWS Well-Architected Framework recommends best practices for ensuring that the infrastructure you deploy is architected in a secure, reliable and efficient way, and optimized for performance and cost. This paper outlines how to align with the cost optimization pillar of the AWS Well-Architected Framework, best practices to establish a cloud financial management practice to reduce AWS costs, and guidance on how AWS customers use a cloud financial management solution such as VMware Tanzu CloudHealth® to execute that practice.

Challenges with cloud cost management

Enterprise IT teams face many challenges with public cloud operations, from ensuring security, reliability and performance to streamlining cloud operations. However, the biggest single challenge for most organizations—and one that still takes many by surprise—is controlling cloud costs.

Ensuring that costs don't spiral out of control is much more difficult in the cloud than it is with an on-premises data center. In the data center, infrastructure hardware has to be purchased ahead of time, so your capacity is limited to the amount of hardware that your IT team purchased. You will never get a shockingly large bill because most costs are upfront capital expenses (CapEx model).

But in the cloud, you obtain compute resources through a subscription service to an infrastructure provider and pay this [operating expense](#) (OpEx) as you go. With this model, you always have enough compute resources to run your workloads, but you run the risk of encountering a big surprise on your cloud bill, as anyone with access to an AWS account can purchase and configure these resources with few limitations.

This is why cloud financial management has become so important in a cloud-first world. CFM is a practice that helps you drive a cost-conscious culture, establish guardrails to meet financial targets, and align cloud operations to broader business goals. Whether you are planning a migration to the cloud or designing new workloads for cloud deployment, CFM will help you maximize business value and achieve financial control.

Cloud financial management should be approached as a discipline built into your organizational DNA—something to be developed over time until cost optimization is recognized as a responsibility of every employee who touches cloud operations.

AWS Well-Architected Framework: A foundation for cloud success

The AWS Well-Architected Framework describes the key concepts, design principles and architectural best practices for building and running workloads in the cloud. The framework is built on six pillars necessary to create a solid foundation for cloud operations. It provides a consistent approach to evaluate architectures and guidance to implement designs that scale with your application needs over time.

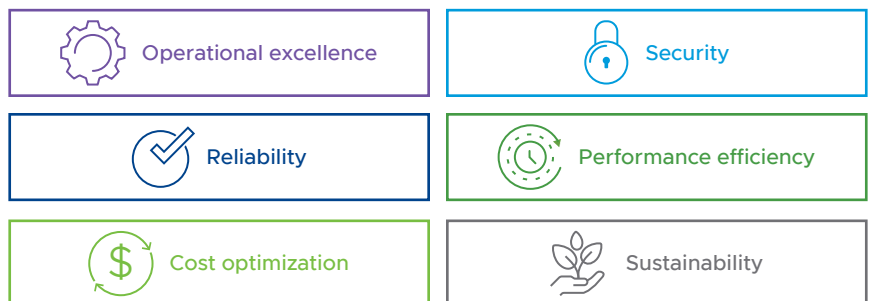


Figure 1: The six pillars of the AWS Well-Architected Framework.

Each of the six pillars of the framework is built on design principles that are applied through best practices. The framework poses a set of foundational questions that reveal how well your architecture aligns with cloud best practices. It also provides guidance for making improvements.

Well-architected pillar no. 5: Cost optimization

This paper is primarily focused on the fifth pillar of the AWS Well-Architected Framework: cost optimization. The goal of cost optimization is to deliver the desired business objectives at the lowest cost. That may sound simple, but it can be difficult to achieve when cloud operations are decentralized across teams, departments and lines of business.

The framework defines five focus areas for your cloud cost optimization journey:

1. Practice cloud financial management – Develop a mature CFM practice, which enables organizations to realize business value and financial success as they optimize costs and scale on AWS.
2. Expenditure and usage awareness – Understand your organization’s cost drivers, which is critical for identifying cost optimization opportunities.
3. Cost-effective resources – Use the correct size, type and number of resources for your workloads, which is key to cost savings. You should also choose a pricing model that’s most appropriate for your anticipated needs.
4. Manage demand and supply resources – Pay only for what you need when you move to the cloud. You can supply resources to match the workload demand at the time they’re needed, eliminating costly and wasteful overprovisioning.
5. Optimize over time – Develop a workload review process to ensure your workloads remain aligned with business objectives and explore new services as needed. Reach maturity by aligning cloud key performance indicators (KPIs) to key business objectives.

Success in these five areas ultimately comes down to having visibility into your environment: understanding your usage patterns, knowing when peaks in demand are expected, and seeing which resources are no longer needed.

In practice: Cost optimization for the cloud

In the following sections, we’ll cover best practices to align with the five focus areas of the cost optimization pillar and demonstrate how AWS customers use the Tanzu CloudHealth platform to establish the visibility and control to ensure they adhere to these principles.

1. Practice cloud financial management

Cloud financial management is an essential part of your cloud strategy—just as important as cloud operations or cloud security and compliance. Organizations that fail to implement CFM often suffer from ballooning cloud costs and a failure to achieve business goals, as development teams are not held accountable when cloud costs affect profitability.

AWS recommends establishing a cost optimization role or team that initiates collaboration between finance and technology, drives a cost-aware culture and processes, and aligns cloud operations to business goals.

The importance of a cloud center of excellence

The cost optimization team is part of an organization's cloud center of excellence (CCoE), and we coach all of our customers to establish this team early on in their cloud journey. A CCoE team is a proven way to drive cultural and operational changes needed to be successful in the cloud by collaborating to design and establish policies and guardrails to stay on track and meet goals.

The CCoE is a cross-functional team that is tasked with supporting and governing the execution of your organization's cloud strategy. The CCoE team will:

- Drive financial accountability and ownership across business groups
- Identify and scale best practices across the organization
- Make business decisions based on accurate ROI analysis
- Understand how all components of a modern cloud environment contribute to total cost of ownership (TCO)

To learn more about how to establish a CCoE team, read the white paper: [The Next Generation of Cloud Management Starts with a Cloud Center of Excellence.](#)

Key phases of cloud financial management

Cloud financial management isn't a one-time exercise. Given the ever-changing nature of cloud infrastructure, the goal of CFM is to continuously optimize and align cloud investments to strategic business initiatives. We define four key phases of CFM, and each successive function indicates a more mature CFM practice:

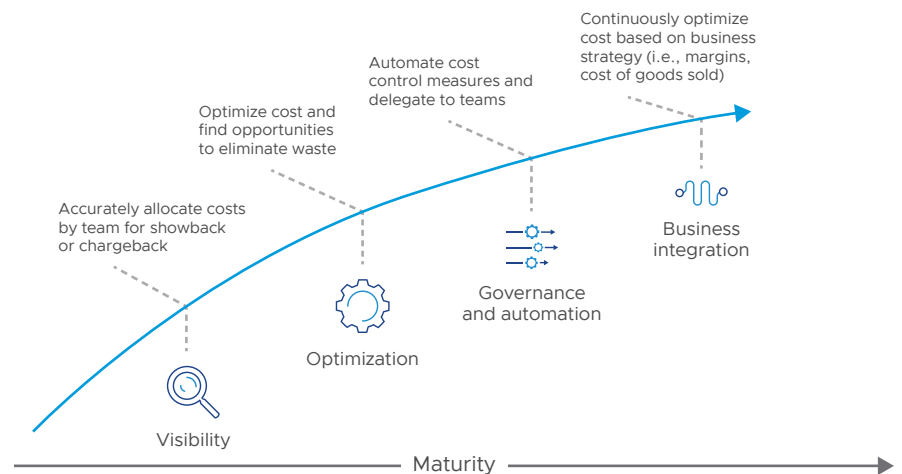


Figure 2: Four key phases of cloud financial management.

- Visibility – Gain granular visibility with custom reports and dashboards broken down by project, team, department or business unit.
- Optimization – Optimize your environment with detailed analysis of resource utilization, including containers and Kubernetes environments.
- Governance and automation – Implement governance policies and automate actions for consistency and control.
- Business integration – Integrate with existing tools and align your cloud strategy with strategic business initiatives.

Achieving visibility is the first step in cloud financial management, and as you progress through each successive step, your business will develop a cost-aware culture and processes. Eventually, you can quantify the business value delivered through cost optimization. When you are able to do that continuously, you have achieved the highest level of CFM maturity.

To learn more about the key functions of cloud financial management and how to improve your CFM practice, read the white paper: [Building a Successful Cloud Financial Management Practice](#).

2. Expenditure and usage awareness

When it comes to monitoring cloud resources and costs, it's easy to see total resource usage and expenditures on your bill. The difficulty comes in segmenting resources and expenditures by groupings that are meaningful to your business. CFM is simply not possible without visibility into your entire organization's cloud spend, and then learning which teams or applications consumed your cloud resources. This visibility is the first step to driving cost accountability.

Additionally, most organizations need to slice and dice their cost, usage and performance data in a number of ways to report different metrics to different stakeholders. For example, the finance department may need a monthly breakdown of costs by product line or shared environment, operations may need a usage breakdown by project or team, and engineering may want a performance breakdown by application role. The ability to segment by different business units, projects or applications is also invaluable for accurate chargeback/showback and understanding which areas are driving significant cloud spend. Having multiple ways to view usage and spending is pivotal to aligning all users to shared goals.

This becomes more challenging as your environment grows. If your business operates in a multi-cloud or hybrid cloud environment, you'll need single-pane reporting to reduce the complexity of cloud management and cost optimization. Even if your environment is restricted to a single public cloud, a centralized cloud management platform can provide easier access to detailed cost analysis than reviewing the output of the native tools offered by your provider of choice.

The Tanzu CloudHealth platform is designed to address these challenges by providing visibility across your entire cloud ecosystem and enables you to continuously analyze and report spend by cost center, drive accountability against budgets, and review recommendations to lower your cloud spend.

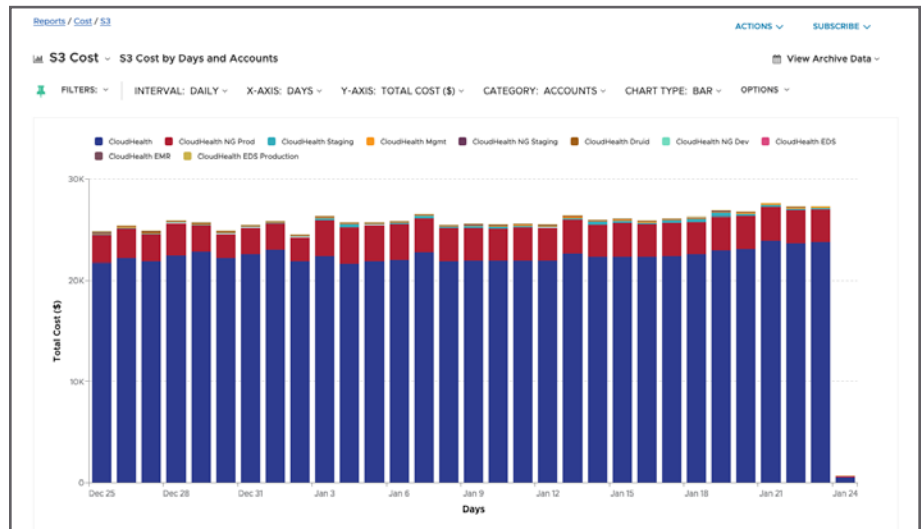


Figure 3: Analyze costs and usage from a centralized location with cost and usage reports in Tanzu CloudHealth.

3. Cost-effective resources

Once an organization has visibility into resource usage and expenditures, the next step in the CFM journey is to optimize costs to ensure that resources are being utilized efficiently. You should always evaluate the cost of new products and services carefully to ensure you select the correct size, type and number of resources for your workloads. Finally, you should choose a pricing model that's most appropriate for your anticipated workloads.

There are two primary ways to keep your resources optimized for cost and performance:

1. Rightsize existing resources
2. Purchase discounts to save money on future resources

Rightsizing recommendations

It's quite common for developers to overprovision resources to ensure their applications function during peaks. When this practice is repeated at scale across hundreds of developers, the amount of wasted spending can be astronomical. It's not unusual for dozens or even hundreds of instances to be overprovisioned and operating at very low utilization. These costs add up quickly and are not returning value to your business.

- Compute rightsizing – With compute, consider CPU, memory, disk and network in/out utilization. Reviewing these trended metrics over time, you can make decisions around reducing the size of your instances without hurting the performance of the application that’s running on it.
- Storage rightsizing – The critical factors to consider with storage rightsizing are capacity, IOPS and throughput. Removing unattached storage is one way to reduce spend and ensure resources aren’t going unused.
- Database rightsizing – Platform-as-a-service (PaaS) solutions, such as relational databases, are used by many developers 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.

It’s important to evaluate your organization’s cloud usage regularly and ask questions about rightsizing. However, without performance monitoring or cloud management tools, it’s difficult to tell when assets are over- or under-provisioned and to prevent this kind of waste on a consistent basis.

Tanzu CloudHealth provides out-of-the-box recommendations to rightsize Amazon Elastic Compute Cloud (EC2) instances (including container nodes) and Amazon Elastic Block Store (EBS) storage volumes to reduce waste and optimize cost. Many organizations set custom thresholds to fine-tune recommendations for their specific needs. This also enables users to take actions on those recommendations directly within the platform, including starting, stopping, rebooting, and deleting compute/ storage, and running AWS Lambda functions.

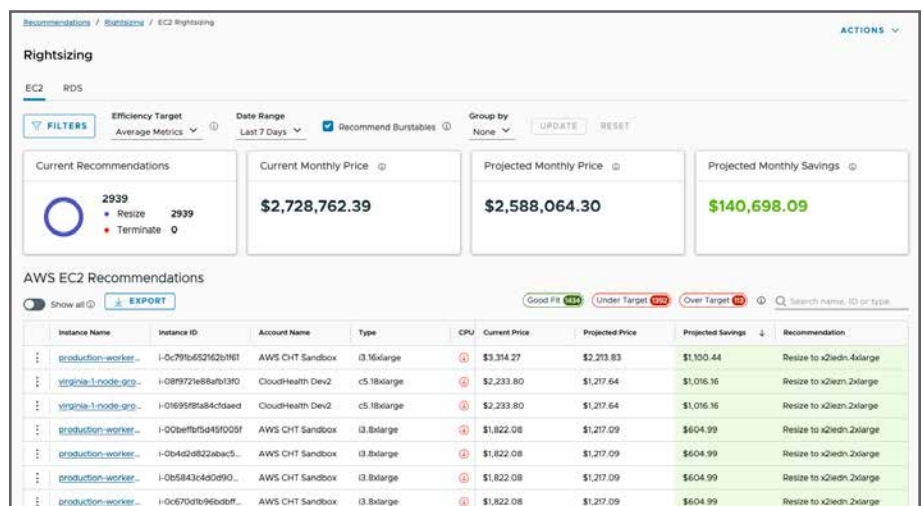


Figure 4: Example of Tanzu CloudHealth rightsizing recommendations.

Reserved Instances and Savings Plans recommendations

Making commitments to consume a specified number of resources or spend a certain amount of money is another way to reduce operating costs in the AWS cloud. AWS offers two very popular types of commitment discounts:

- Reserved Instances – Pay up to 72 percent off on-demand prices for EC2, Amazon Relational Database Service (RDS), Amazon ElastiCache, Amazon Redshift, Amazon DynamoDB, and Amazon Elasticsearch in exchange for an upfront commitment to use those resources over a period of one or three years.
- Savings Plans – Pay up to 72 percent off on-demand prices for EC2, Lambda and AWS Fargate in exchange for an upfront commitment to spend a certain amount of money per hour over the course of one or three years.

With savings this great, commitment discounts create the opportunity for significant cost reduction. However, they can also introduce additional complexity when it comes to cloud financial management.

Managing Reserved Instances and Savings Plans

AWS customers often find that it's simply not efficient to try to optimize a large and dynamic AWS environment using complex spreadsheets.

AWS provides native tools to help businesses identify opportunities to purchase Reserved Instances and Savings Plans, but these tools often fall short in terms of functionality and can rack up unexpected costs. As businesses incorporate these pricing options into their strategy, many find that they need more than what these tools can provide, including where and how the discount programs are being used. This is particularly true in cases where excess discounts float across regions and between instance families, or get dispersed throughout consolidated accounts, as well as with organizations that operate in multi-cloud or hybrid cloud environments and need granular visibility and discount recommendations across clouds.

It's also important to know when Reserved Instances and Savings Plans are not being fully utilized. Although you can resell unused Reserved Instances through the AWS Marketplace, that's not the case with Savings Plans. Once you've made a commitment to spend a certain amount each hour over the next one or three years, you're tied to that commitment for the lifetime of the Savings Plan. With this in mind, you need to stay on top of how your discounts are being utilized at all times.

Our customers use the Tanzu CloudHealth platform to track which accounts benefit from discounted prices and calculate net compute costs down to business units, projects, teams, individuals or logical business groupings. The platform can also be configured to alert administrators to unused discounts, while the Convertible RI Exchanger can find the mathematically optimal use of the discount programs.

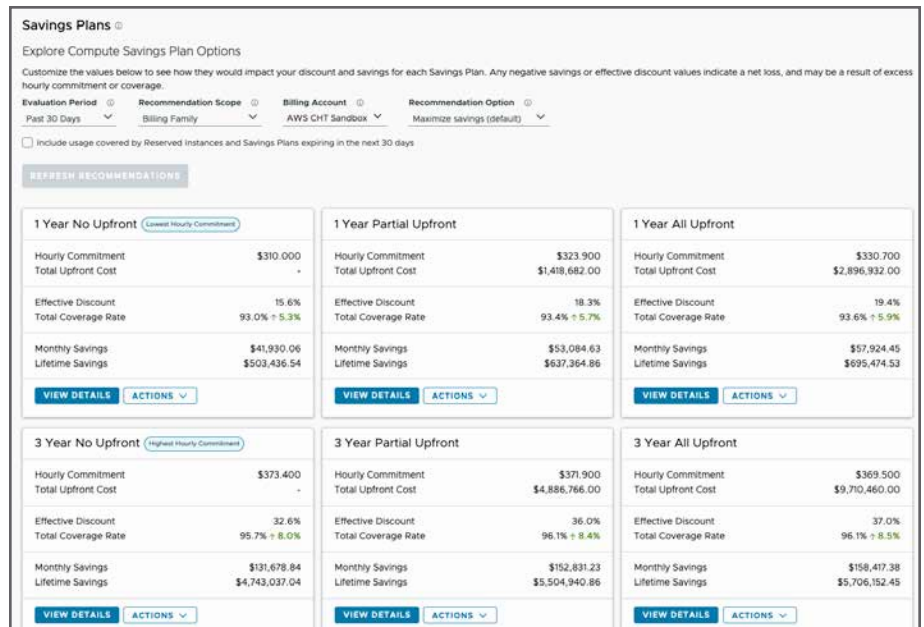


Figure 5: Tanzu CloudHealth price breakdowns of various AWS Savings Plans.

Tanzu CloudHealth enables organizations to quickly optimize resource selection and usage, and manage AWS Reserved Instances and Savings Plans, so you can be sure you're always using the most cost-effective resources.

4. Manage demand and supply resources

Rightsizing existing resources and optimizing usage can be extremely helpful for cost control, but to supply resources as needed at scale without blowing your budgets, rule-based policies to govern cloud operations are critical. Policy-driven automation is a great way to maintain cost control in a decentralized cloud environment.

Policy-driven governance and automation

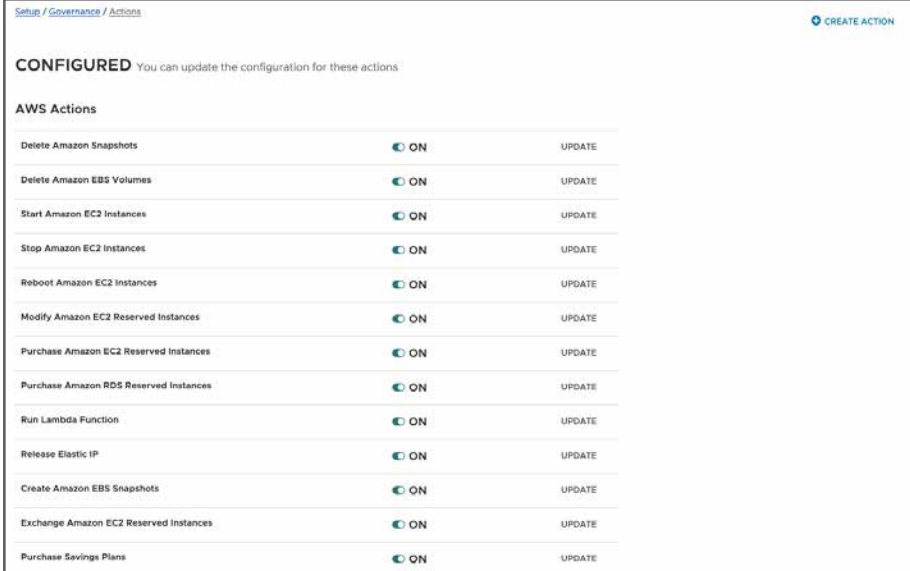
The basic concept of policy-driven automation is that a system administrator creates a rule—also known as the policy—and configures your cloud management platform to monitor your cloud environments for compliance with the rule. If a violation of the rule occurs, the platform will notify the authorized user(s) or perform a predetermined action.

Examples of popular rules and actions include:

- If month-to-date spend is projected to exceed 100 percent of the budget, send a notification to the budget owner
- If costs (per resource type, tag, project, environment, and so on) increase by more than 10 percent in a week, send a notification to the budget owner
- If a resource is deployed above a certain capacity, trigger a function to stop the resource and initiate an approval workflow
- Automatically delete unused EBS volumes if no activity is detected for 10 or more days
- If a resource is deployed outside a pre-authorized region, trigger a function to terminate the instance and send a notification to the owner's team leader
- If an EC2 instance has unauthorized open ports, initiate a function to terminate the resource and notify system administrator

There are many different uses for policy-driven automation. In addition to cloud cost management, customers can use it to automate the optimization of their cloud environments, ensure end-user experience is not impacted by under-provisioned resources, and prevent configuration errors from creating security and compliance risks.

The dynamic policy engine of Tanzu CloudHealth ensures teams comply with utilization and provisioning guidelines set by your company. Your environment is monitored for usage violations and stakeholders are alerted proactively if provisioned resources violate those guidelines. Automation policies can also be used to take immediate action to ensure compliance.



Setup / Governance / Actions CREATE ACTION

CONFIGURED You can update the configuration for these actions

AWS Actions

Delete Amazon Snapshots	ON	UPDATE
Delete Amazon EBS Volumes	ON	UPDATE
Start Amazon EC2 Instances	ON	UPDATE
Stop Amazon EC2 Instances	ON	UPDATE
Reboot Amazon EC2 Instances	ON	UPDATE
Modify Amazon EC2 Reserved Instances	ON	UPDATE
Purchase Amazon EC2 Reserved Instances	ON	UPDATE
Purchase Amazon RDS Reserved Instances	ON	UPDATE
Run Lambda Function	ON	UPDATE
Release Elastic IP	ON	UPDATE
Create Amazon EBS Snapshots	ON	UPDATE
Exchange Amazon EC2 Reserved Instances	ON	UPDATE
Purchase Savings Plans	ON	UPDATE

Figure 6: Example of a configured automated action.

With these tools, Tanzu CloudHealth customers can maintain greater control over resource provisioning and costs.

5. Optimize over time

The final best practice in the AWS cost optimization pillar is optimizing over time. Cloud financial management is not a one-and-done event. Successful CFM includes integrating a workload review process into your standard operations to ensure your cloud environment remains optimized. You should also continue to explore new and updated products, services and discount options, as AWS continues to release more innovative solutions that can make optimization more effective.

But achieving CFM maturity means more than continuously optimizing infrastructure and costs. Cloud financial managers need to understand what usage and cost metrics mean in the context of the business. For example, suppose the cost of running a particular application in AWS goes up by 20 percent. If you're only measuring costs, this looks like bad news. But if your revenue has simultaneously increased by 50 percent, then that extra spend was well worth it. For this reason, cloud KPIs should be aligned to business goals such as increasing net profits.

Learn more

To find out how Tanzu CloudHealth can help with your cloud financial management, [sign up for a free trial](#).

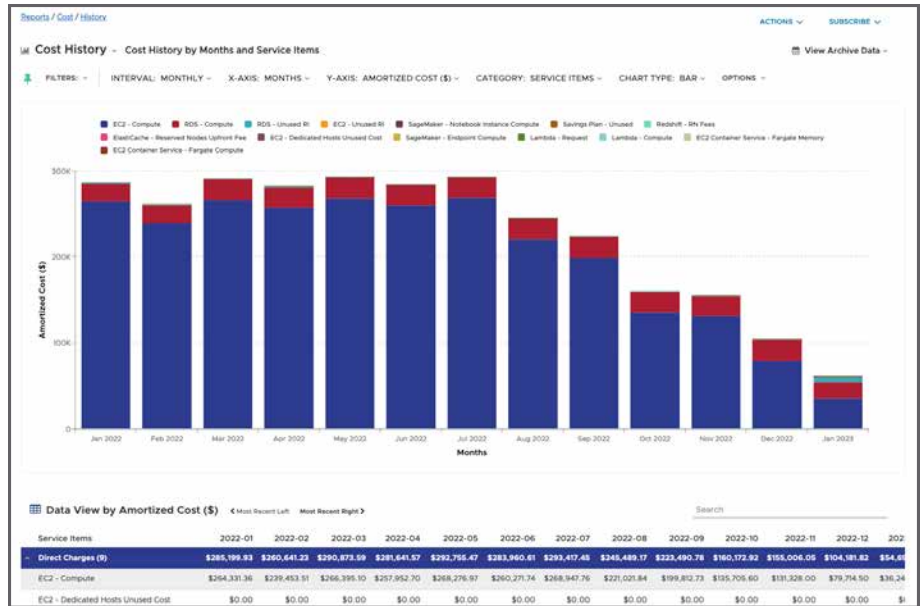


Figure 7: Example of how Tanzu CloudHealth can help to optimize spend.

Conclusion

A mature cloud financial management practice has become essential for organizations operating in the cloud, as cloud cost management surfaces new and exciting challenges. The AWS Well-Architected Framework can help your organization architect a secure, reliable, and cost-efficient environment to support all your workloads. The framework’s fifth pillar identifies five focus areas for achieving cost optimization.

Tanzu CloudHealth offers capabilities that align closely with the cost optimization pillar, enabling your organization to reach cloud financial management maturity. With Tanzu CloudHealth, organizations can more easily utilize the design principles and best practices outlined in the [AWS Well-Architected Framework](#).

To get a sense of where your business currently stands on the maturity curve for cloud financial management, [take our five-minute assessment](#).

