



How to Gain Control of Your Cloud Spend

An Introduction to Cloud Financial Management

Introduction

The public cloud is over a decade old now, and its basic benefits—rapid scalability, flexibility, and consumption-based pricing—are well known. Over that time, we’ve seen how cloud services, when deployed smartly and utilized strategically, can be revolutionary for businesses.

For most organizations, it’s no longer a question of whether public cloud makes sense for them; it’s a question of when, how, and what they can do to accelerate adoption as quickly as possible.

But as we’ve learned over the years, tracking cloud spend and directing funds to get the most business value isn’t simple. The cloud operating model is different, and it requires a new way of thinking about IT procurement and cost management.

This eBook will introduce the concept of cloud financial management and explain how you can use it to gain control over your cloud spend, improve your forecasting efforts, and make better investment decisions.

“You wouldn’t go to cloud without thinking about security. So why would you go to cloud without thinking about finance?”

Nathan Besh,
Sr Director, Product Management and
Technical Evangelism

What is Cloud Financial Management?

Although the cloud operating model may be different, your job as an IT finance leader hasn't changed. You're still responsible for forecasting expenses, setting budgets, tracking spend, and optimizing investments to ensure your organization is getting the best value for its money.

How do you do that in a cloud environment? When it comes to financial management, the cloud introduces several challenges.

First, IT procurement is no longer centralized. Business units, individual engineers, and Infrastructure as Code (IaC) can provision cloud resources, usually without the need for prior approval or authorization. While most organizations have now moved to consolidated billing with their cloud vendors, the removal of gates for procuring infrastructure means IT finance must adapt to more unpredictable spending.

“**67%** of enterprises say CFM helped them grow revenue, and 64% say it helped them increase profitability.”



Second, without adequate policies and common practices in place, assigning cloud resource usage from the consolidated bills to specific applications, business units, or departments is nearly impossible. Cloud providers allow you to create new virtual machines and other resource types, expanding computing and storage resources as needed to handle new workload requirements, without requiring meaningful naming conventions.

Third, the inherent nature of cloud provisioning creates a complex billing structure. Cloud providers offer hundreds of thousands of SKUs (stock-keeping units) and bill for services based on usage at the resource level. While each cloud vendor bills with a different granularity, in the case of AWS the most common scenario sees one billing line item for every resource hour incurred. This results in monthly bills containing millions (sometimes billions) of rows, making cloud billing a big data concern.

Cloud Financial Management (CFM) is about gaining control over your cloud spending and holding delivery teams financially accountable. Fundamentally, CFM is a business discipline—a set of processes and systems designed to help you increase visibility into the cloud resources your organization is using so you can allocate their costs to the proper business units, forecast spending, reduce and/or optimize cost, and make better decisions about where to invest your organization's finances.

How Does Cloud Financial Management Help You Gain Control of Cloud Spend?

It's natural to equate financial management in the cloud with reducing costs. After all, no one wants to waste money, and provisioning cloud resources is so easy to do, it's not uncommon to lose track of what has been spent. While eliminating wasteful spending is a goal of CFM, it's not the only one. It's worth noting that there are two parts to Cloud Financial Management: cost management and optimization. For you to really take control of your cloud spending—and direct your investments to where they will generate the most value for your organization—you'll need insight into both areas.

Cost Management

Cloud Cost Management is all about assigning ownership and providing cost visibility. You need to understand what resources your organization is consuming and who is actually responsible for them before you can make recommendations for change. Across the public clouds, assigning ownership is primarily done using a combination of two constructs: accounts (or equivalent) and tags.



Accounts, Subscriptions, and Projects

The first logical way to organize your cloud spend is through the use of accounts -that is, if you're using Amazon Web Services (AWS). If you're using Microsoft Azure, the equivalent is called subscriptions. If you're using the Google Cloud Platform (GCP), the equivalent is projects. Cloud providers allow you to create as many of these as you need. Whenever anyone provisions a cloud resource, such as a virtual machine (VM) or volume, they must decide which account is most suitable for hosting it. This enables you to define an overarching structure for your cloud environment centered around logical conventions, naming and grouping accounts, subscriptions, or projects for different areas of your business (for example, by business unit).

In AWS, you can build this structure around one or a handful of payer accounts (now known as "management accounts") with each backed by numerous member accounts (where the usage happens). For example, you could create a management account for each of your main business units. Then, you could create individual member accounts that are dedicated to product areas within the business unit. Enterprises are increasingly shifting to a single management account model to maximize Reserved Instance and Savings Plan benefits. Both Microsoft Azure and Google Cloud Platform provide similar grouping capabilities through subscriptions and projects respectively.

It is worth noting that every resource belongs to an account by necessity, and therefore it's the most foolproof way of assigning top-level ownership. If you need to capture more detailed information about what is driving your cloud costs, then you'll also need to consider leveraging resource tags.



Tags

Resource tags are like labels you put on physical assets. They consist of a customer-defined key-value pair. Tags can provide crucial information about any cloud resource —such as its name, environment, service, application, department, division, and much more. You can consider tags an important type of metadata that allows you to apply detailed information to the resources (i.e., the compute, storage, and network services) your organization is consuming, and in turn, drive cost management.

Using tags, you can be as granular as you want in describing your resources. For instance, AWS allows 50 user-created tags to be attached to each resource, as does Microsoft Azure. GCP calls these “labels” instead of “tags” and allows up to 64 per resource.

Be aware that tagging is case sensitive, and spelling matters. Because you have the ability to attach so many attributes to each resource, it’s easy to get carried away with tagging if you’re not careful. In addition, not every resource type supports tags, so be sure to check your provider’s limitations.

Your best approach is to start small. Use a tagging strategy that is simple and covers the most important reporting items, and then expand it over time if necessary.

Once your tagging and account strategy is in place, you’ll be able to leverage attributes within detailed billing to allocate costs, then use showback or chargeback processes to hold business groups accountable.



Showback

As an IT finance leader, it's important to set financial metrics for cloud spend and to track performance against them. Each entity in your organization, depending on how you structure your financial reporting, can then be held accountable for managing its spending and not exceeding budgets.

Showback is the official process of allocating your cloud bill to the different parts of your organization, providing budget-owning entities with a view of their financial performance. The work done to tag resources and thoughtfully group accounts provide the foundations for this showback. Additional considerations will include how to allocate out commitment purchases, such as Reserved Instances and Savings Plans, how to pass on/not pass on custom pricing agreements, and what to do with shared costs such as enterprise support charges.



Chargeback

Chargeback is largely the same as showback—both refer to the official process of reporting back cloud cost figures for the purpose of managing budgets and holding individual areas of the organization accountable for spending.

There is, however, one significant difference between showback and chargeback. With showback, you are simply reporting financial performance for management purposes; the actual incurred cloud costs are not applied to the profit and loss (P&L) statement of the individual business units. On the other hand, chargeback has a direct effect on each business unit's financial results. Whatever costs are allocated to the business entity are applied to that unit's P&L statement.

For More Information

[➔ How Allocation and Visibility Drive Cost Accountability](#)

Optimization

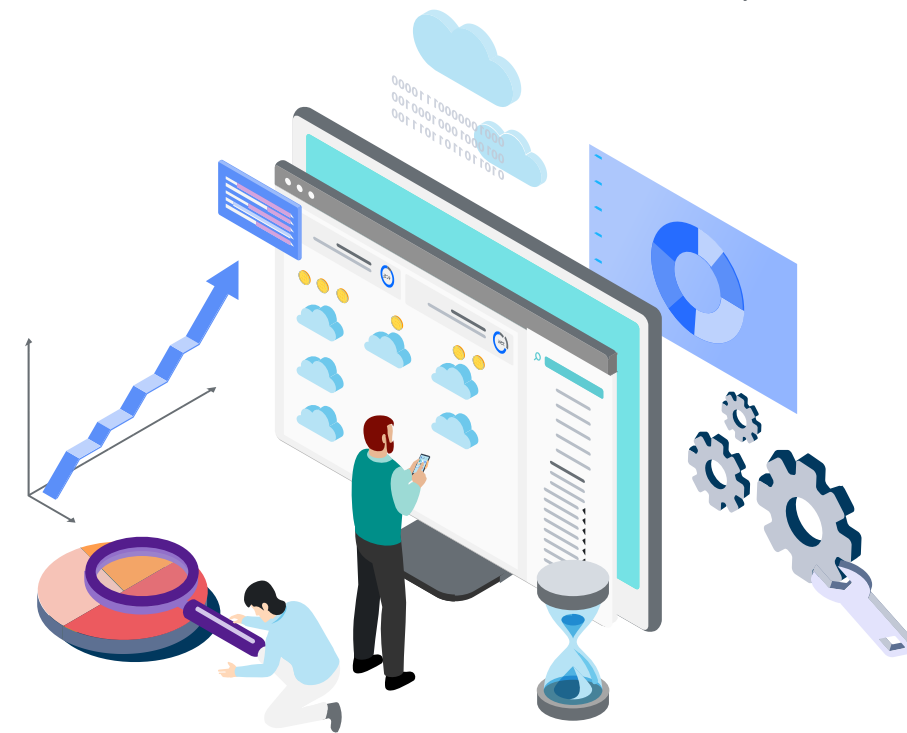
Cost management is the crucial first step to Cloud Financial Management that brings visibility, helping you understand the cloud resources your organization is consuming and who is consuming them. But this is only the first step.

The next step is to optimize: to find ways, by analyzing cost and usage data, to make better provisioning choices for the resources you're already paying for (usage optimization), and to reduce the hourly rates paid through purchasing commitments (financial optimization).



Usage Optimization

One of the most important things to know about moving to public cloud is that the consumption-based billing model means that you pay for what you provision, regardless of how well these resources are utilized or even if they are used at all. Given that the cloud vendors meter their billing by the second (AWS and GCP) or minute (Azure), provisioning decisions have an immediate and ongoing impact.



7

Usage optimization is focused on surfacing insights into how well matched your cloud resources are to their underlying workloads and identifying practical provisioning options that will reduce costs. Broadly speaking, there are two categories of usage optimization: rightsizing and idle resource discovery.

Rightsizing refers to the processes and techniques for making intelligent trade-offs between performance and cost across cloud footprints. It requires an understanding of what your current cloud resources are costing you, to what degree each is being utilized (across metrics, such as CPU and memory, that must be gathered from monitoring solutions), and identifying alternative options (such as reduced size or different type) that will meet the needs of the underlying workload. A simple rightsizing scenario might occur where an engineer provisions a large VM (hence high hourly rates) for a task that requires minimal CPU or memory. By rightsizing to a smaller VM, costs can be reduced with minimal performance risk. There are many different rightsizing scenarios that can occur and being successful in getting engineers to act when running cloud at scale requires specialist tools and dedicated strategies.

The second area where cloud spend can be wasted is on entirely idle or orphaned resources. It's common for teams to spin up new cloud resources to support short-term projects, and when the projects are complete, the teams move on to new projects. This leaves resources sitting idle, supporting no workloads at the ongoing expense of the organization. Given how easy it is to provision cloud infrastructure, it's not surprising that engineers will forget to turn some things off. Likewise, there are many scenarios where cloud resources unintentionally become orphaned (e.g., volumes that become unattached or load balancers with no workers). Idle resource discovery is the process of analyzing cloud resources, reviewing configuration and utilization metrics to identify those that are not being used so that they can be decommissioned, reducing your overall cloud spend.



Financial Optimization

Financial optimization is about improving your cost efficiency by taking advantage of financial instruments provided by the cloud vendors. By committing to levels of specific types of usage, the vendors offer attractive, lower hourly rates. Financial optimization has the benefit of not requiring engineering effort to receive efficiency gains. However, care must be taken to not over-commit, as any non-used commitment hours become a waste. Achieving the goals of maximizing commitment coverage while minimizing commitment waste requires detailed analysis and ongoing management.



Reserved Instances

Offered by AWS and Azure, Reserved Instances (RIs) are the original commitment instrument. RIs can be purchased for a range of cloud service usage such as VMs (AWS Elastic Computing (AWS EC2) or Azure Compute) and managed databases (AWS RDS or Azure SQL Database). You can think of RIs as discount coupons; you purchase a set of coupons that get applied to your usage every hour. Every hour, the cloud vendor resolves which of your usage qualifies for the coupons and decides which specific resource usage will get the benefit (important: you can't purchase RIs for specific resources; rather RIs "float around" based on the vendor's algorithms). Instead of being charged full on-demand fees (say, \$1 for the hour), the usage will receive the RI rate (say, \$0.70).

RIs are typically purchased based on three attributes: location, operating system, and VM type/size. For example, for AWS you could choose to purchase 10 m5. large EC2 RIs in US-East-1 for Linux. If any RIs go unused for a given hour (i.e., there wasn't adequate usage to consume them), then they become waste. When purchasing a reservation, you choose a length (one year or three years) and payment model (all upfront, partial upfront, or no upfront) with the savings rate dependent on your choice. Given the complexity around savings rates and the need to analyze hourly data across a wide combination of usage, we advise using a purpose-built recommendation engine that allows you to create a plan matched to your organization.

Savings Plans

In recent years AWS has simplified RI management by adding flexibility with innovations like regional benefit and instance size flexibility. This has culminated in the launch by AWS of a completely new commitment instrument, Savings Plans. You can think of Savings Plans in largely the same way as RIs: discount coupons that get applied hour-by-hour and with identical payment options and saving rates. Savings Plans are superior to RIs in that individual commitments have a higher level of abstraction, covering a broader set of underlying usage. The most obvious difference you'll notice is that commitments are made in dollars (\$) rather than instance count.

There are two categories of Savings Plans. The most flexible type is known as a Compute Savings Plan. With this type, each commitment is restricted only to compute in a specific AWS region – it will cover all instance families/sizes and operating systems. Compute Savings Plans will also float across different cloud services such as EC2 and Lambda (AWS is continuing to add supported services). The second category of Savings Plans are the service-specific ones, such as “EC2 Instance Savings Plans.” These commitments are purchased specific to the region and instance family, but they will float between sizes and operating systems. EC2 Instance Savings Plans have the same savings rates as Standard RIs, while Compute Savings plans have the lower savings rates that match Convertible RIs.

Committed Use Discounts

If you're a GCP customer, you can take advantage of Committed Use Discounts (CUD). This cost-saving instrument is the equivalent of the RIs offered by the other cloud vendors. CUDs are available only in a no-upfront payment type but can be made for one-year or three-years. The biggest difference you'll notice with CUDs when compared to RIs is that you are reserving specific amounts of memory and CPU rather than instances, matching the general billing model of GCP.



For More Information

[➔ Mastering the Fundamentals of AWS Cost Efficiency](#)

What If You Use Containers?

As you begin to deploy more advanced technologies like containers, cloud cost allocation becomes more challenging.

Containerization makes it easier to package and run software applications in different computing environments. Running containers at scale is done by leveraging an orchestration tool, typically Kubernetes (K8s). With operational benefits around agility and resource efficiency, it's no surprise that K8s deployments are steadily growing as a percentage of public cloud spend.

From a cost allocation perspective, it's worth noting that each K8s cluster is usually a multi-tenant entity, with underlying virtual machines and volumes being shared across applications and teams. This is why some people refer to these deployments as a cloud-inside-a-cloud, and describe a "black-box" problem in attempting to understand who is responsible for consuming the associated cloud resources.

The cloud vendors offer no native billing construct to surface the cost of each cluster or means to split their cost up so they can be allocated back to the business. It's highly recommended to take advantage of specialist solutions, such as Cloudability, that can automatically map resources to their cluster, calculate total cluster cost, and then apply rules based on usage patterns to split these costs up and allocate them out via K8s constructs such as Namespace and Labels.

What About SaaS?

No discussion on cloud financial management would be complete without a word on Software-as-a-Service (SaaS). According to Gartner, SaaS is the largest market segment of public cloud services, expected to account for over \$122 billion in 2021.

For most organizations, SaaS adoption is growing, but is it being used efficiently? Are funds paying for services that are not being used? Is money being spent on duplicate applications without anyone in the organization realizing it? Could money be better spent elsewhere?

As an IT finance specialist, your job is to answer those questions. SaaS presents similar challenges for you as other public cloud services do, and without proper systems in place, you have limited visibility to your organization's full IT footprint, especially shadow IT, and that can inhibit your ability to analyze usage and make informed investment decisions.

Resources to Help

Cloud Financial Management (CFM) as a business discipline continues to evolve. Developing processes and systems takes focus and determination, and just getting started can be the most challenging part.

Fortunately, you're not alone. Organizations like the FinOps Foundation, with over 2,700 members representing more than 1,200 companies, exist to provide best practices, education, and support for the ongoing maturity of financial operations (FinOps) and cloud financial management. In addition, the TBM Council, a nonprofit organization dedicated to advancing the discipline of technology business management (TBM), has added specific foundational pillars to its framework to account for cloud costs. Tools like Cloudability are available to help you adopt and execute CFM principles.

For more information on FinOps and the discipline of Cloud Financial Management, check out:

[FinOps: A New Approach to Cloud Financial Management.](#)



About Apptio

Apptio is the leading provider of cloud-based Technology Business Management (TBM) software that helps CIOs manage the business of IT. Apptio's suite of applications uses business analytics to provide facts and insights about technology cost, value, and quality so IT leaders can make faster data-driven decisions and align technology spend to business outcomes. Hundreds of customers, including more than a third of the Fortune 100, choose Apptio as their business system of record for IT. For more information, visit the Apptio website at www.apptio.com.

To learn more about how Apptio Cloudability can help you gain control of your cloud spend, go to www.apptio.com/products/cloudability/ where you can access helpful resources and start a free trial.

