



AWS Cost Optimization: Go from Zero to Hero

Intro

Let's face it - cost optimization on the cloud is hard. Only 15% of those who implement FinOps strategies would currently describe their organization as having a "mature and evolving" FinOps practice. That leaves 75% of the world's public cloud users wondering how to encourage and evolve their cost optimization strategies to see greater impact on their bottom line. What does this mean for today's DevOps teams, who need to be able to build freely on the cloud to keep up the pace of change, but who also need to be held accountable to spiraling cloud costs?

In this white paper, we'll cover:

- ▶ **Understanding the problem:** From idle resources to poor optimization, why do cloud costs get out of control? We'll list the main struggles for today's organizations.
- ▶ **Overcoming the challenges:** Top tips for reducing costs and implementing cloud cost optimization best-practices, plus practical tools and products that can reduce the squeeze.
- ▶ **Implementing automation:** How the latest technology uses AI to automate cloud provisioning, make smart real-time decisions and ultimately save you serious spend.

What are the major cloud cost challenges CIOs and CTOs are tackling in 2022?

Read our report to find out!

[Get the Report >](#)

Developing a FinOps team, identifying key cloud stakeholders, implementing governance processes, and putting in a CFM tool... will go a long way to ensure cloud costs are being tracked and managed properly.

Understanding the Problem of Spiraling AWS Cloud Costs

With so many people citing cost savings as the main reason to move infrastructure to the cloud, it can be confusing to understand why cloud spend is such a huge issue for today's businesses. The truth is, while cost is one of the top two drivers for cloud migration for 32% of organizations, once you arrive on the cloud and start to scale and mature your business growth, ROI tends to decline and even plateau. 68% of IT decision makers say that their cloud infrastructure has become more expensive in the past year, and a lack of best-practices and unnecessary cloud costs could be leaving an untapped \$500B on the proverbial table.

AWS calls out 5 best-practices for cost optimization of cloud architecture. If you're failing at even just one or two of these, you can easily see that reflected on your monthly cloud invoice:

- ▶ **Cloud Financial management:** The cloud's dynamic nature makes keeping track of costs difficult without the right processes and procedures in place. Developing a FinOps team, identifying key cloud stakeholders, implementing governance processes, and putting a CFM tool in place for managing Reserved Instances or Auto Scaling will go a long way to ensure cloud costs are being tracked and managed properly.
- ▶ **Expenditure and usage visibility:** The flexibility of the cloud enables various teams across the organization to utilize cloud services simultaneously, making it difficult to gain holistic visibility into usage. Poor integration between cloud systems, and a lack of alignment between the cloud tools and technologies that teams deploy often results in duplication of data, efforts, and spending across multiple teams. Tagging resources, setting up monitoring tools, and using tools like AWS Cost Explorer can help provide more visibility into spend.

AWS offers a variety of cost efficient pricing models to help their customers get more bang for their buck.



▶ **Cost Effective Resources:** AWS offers a variety of cost efficient pricing models to help their customers get more bang for their buck. Reserved Instances offer up to 70% in discounts off on-demand EC2 usage, Saving Plans offer a smaller discount but provides greater flexibility if you want to make changes to EC2 commitments, and Spot Instances can provide up to 90% in discounts, but only work for short-term usage in limited circumstances. You also might want to consider your options for reduced redundancy storage. We'll look into all of these models in more detail later on in this guide.

▶ **Managing Demand and Supply Resources:** As engineers are tied to fast development, they often over-provision on resources, so that they are not left unstuck if they need to scale quickly or unexpectedly to meet demand. This cloud waste has a serious impact on your bottom line, and it's a common problem. In fact, 66% of organizations admit to over-provisioning last year. On the other hand, under-provisioning is also a problem, where developers are left without the resources that they need, often because of cloud cost optimization strategies that are attempting to address spiraling cloud costs.

▶ **Optimizing over time:** As your requirements change, failure to remove idle or unused resources, consider cloud waste, and implement new features and resource types can come back to bite you in an expensive kind of way. Most organizations admit to having idle/unused resources in their system, which can be anything from resources that run full time even though they aren't in use for more than a few hours, or forgotten instances that were used for development or testing environments and weren't shut down after use. Sometimes idle or unused resources are a mistake, for example when education is poor around managing a cloud environment, or when organizations don't schedule their VMs ahead of time and let them run around the clock. Other times it can be intentional, because DevOps over-provision when they don't want to be caught without the resources they need.



How to Overcome Top Cost Optimization Challenges

Enough doom and gloom, it's depressing. Luckily, as these problems are almost ubiquitous in the cloud community, there are also some great fixes that can translate into serious savings.

First, find ways to adapt your architecture to consider costs wherever that's possible. This is a "fun" balancing act, as you'll probably have anyone from security stakeholders to product teams breathing down your neck about the need to simultaneously hold back and measure the risk while you also forge forward and embrace the pace of change. For example, you might want to work on network configurations or regions to get a better price on data transfer. EC2 costs 20% less in Oregon than it does in Northern California, so this could be a smart move from a cost-savings perspective. Wait, though. Will this change impact the availability and performance of your applications? You need to have that conversation before you hit the launch button.

Sometimes the stars will align, and everyone's goals will be on the same page. Find a way to make the process less manual and more automated - and you'll get a smile from everyone involved.

Next, ownership is absolutely critical when you're considering Cloud Financial Management, which is why the best practice of setting up a Cloud Center of Excellence (CCoE) has emerged.

According to Gartner, "To ensure cloud adoption success, organizations must have the right skills and structure in place. The optimal way to achieve this is by setting up a centralized cloud center of excellence (CCOE). A CCOE is a centralized governance function for the organization and acts in a consultative role for central IT, business-unit IT and cloud service consumers in the business. A CCOE is key to driving cloud-enabled IT transformation."

Striking a Balance Between Cost & Performance

Find a way to make the process less manual and more automated!

Attacking the Struggle of Ownership & Accountability

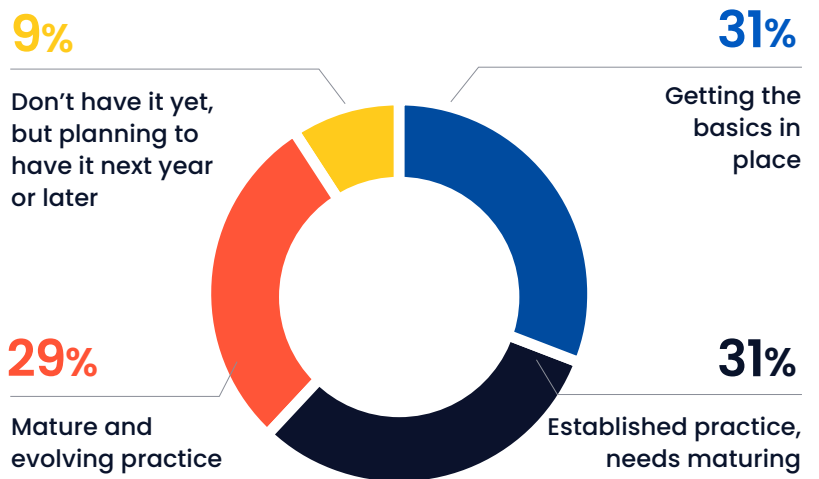
A CCoE will establish who is responsible for the cloud within the business, and ensure there is clear ownership over cloud resources.

A CCoE will establish who is responsible for the cloud within the business, and ensure there is clear ownership over cloud resources, however you want that to work in your business. You can also incentivize engineers to take control over their processes, implementing cost-efficient strategies and technologies as part of the build from day one, and turning off unused resources.

Providing education in order to incentivize engineers to take ownership of their resources is another critical piece of the puzzle. Empower them to play an active role in the FinOps process and remind them that FinOps doesn't mean you want to take money from their budget, but rather reallocate it where it provides strategic value. That means that many of them might actually have more moolah to play with, if they can prove it's needed with visibility into their existing usage and optimization.

For more information on setting up FinOps in your organization, check out our FinOps cheat sheet!

Get the Cheat Sheet >



FinOps Maturity According to CIOs and CTOs

Using AWS's Existing Services for Cost Optimization

Of course, engineers can't do it alone, and why should they have to when there are many specialized products around to make it easier? Business intelligence tools, auto scaling, anomaly detection and monitoring tools are all offered natively in AWS, and can offer insights into cloud usage such as detecting unused or idle resources, zeroing in on orphaned EBS volumes, or which EC2 instances are racking up the biggest expense. [AWS Cost Explorer](#) is a great start, helping you to visualize and manage your AWS spend over time with default reports for both services and resources.

Automation is also a cost optimizer's best friend!

A Word on Private Pricing

Yet currently only 65% of those at the 2 year mark with AWS are utilizing AWS Cost Explorer at all, and even those who do will realize it's value only a visualization tool, Indeed, it doesn't perform any actual actions itself - leaving those up to the engineer.

Automation is also a cost optimizer's best friend, and can reduce the guessing game of purchasing instances and commitments, as well as helping to autoscale storage volumes behind the scenes, or right sizing compute resources. Let's take AWS Auto Scaling as an example. This native AWS tool will ensure you always have the right number of EC2 instances to handle the application load. You can specify a minimum and maximum number of instances in each group, and then set smart policies for launching or shutting down instances as demand changes.

You may have heard of the AWS Enterprise Discount Program, which has now been updated to PPTS - Private Pricing Term Sheet. (Catchy, eh?) Here's how it works.

If you're an enterprise AWS customer who can commit to large cloud spend, especially over a long period of time, AWS will provide a discount in return for a pre-paid, annual consumption commitment. There is a minimum commitment of \$500,000 per year, and the discounts are negotiated with AWS on a case by case basis, but generally speaking, it gets more exciting the longer your commitment term and the higher your spend. Don't forget, you need to be signed up to Enterprise Support to qualify.

Of course, you'll need to be careful about overcommitting to cloud usage and spending more than you would have needed to, and at the same time, ensure you aren't under-committing and missing out on further discounts. It's another one of those balancing acts we know and love.

Thinking About Pricing Models

It's therefore important to consider what you're going to need ahead of time. RIs can't be canceled, and generally speaking can't be modified.

If you have relevant stateless, fault-tolerant or flexible applications - you could enjoy a 90% discount on regular prices.

Still using too many On-Demand instances? Then you're probably seeing it in your cloud bill.

Here are some of the main ways to cut costs through AWS's pre-defined models.

- ▶ **Reserved Instances:** This is a simple billing discount on your EC2 usage costs. You'll commit to a contract term that's either one or three years, and then be billed for each hour regardless of whether you use the instance. Standard RI savings are on average 40% for a one year commitment, and increase to 60% for a three year term. It's therefore important to consider what you're going to need ahead of time. RIs can't be canceled, and generally speaking can't be modified, as you'll always be left with certain attributes. While they can be sold on the EC2 RI Marketplace if you no longer need them, usually the sale price is well below market rates.
- ▶ **Savings Plans:** There are two different kinds of savings plans, one for compute and the other for EC2 Instances. Compute is a lot more flexible, and can be utilized for EC2, Lambda and Fargate, across family, size, region, OS or tenancy. For EC2 Instance Savings Plans, you'll need to commit to a single family type in one region. Simply put, commit to a particular instance type in a specific AWS region for one year or three years, and get a discount off your On-Demand spend of up to 72% on EC2 and as much as 66% on the compute savings plan.
- ▶ **Spot Instances:** Spot Instances are the most inexpensive way to buy EC2 capacity, but there are limited use cases for most businesses, as AWS can request the capacity back with two minutes notice. If you have relevant stateless, fault-tolerant or flexible applications - you could enjoy a 90% discount on regular prices. Test and development workloads are a good choice, or big data training models where a hibernation at short notice isn't a problem. Remember that workloads will need to be configured in order to leverage them, so you'll need to take into account the extra engineering effort.

AWS EC2 Pricing Models



ON-DEMAND INSTANCES

No Commitment

High Flexibility

No Upfront Payments

Easy to Work With

Most Expensive Option



SPOT INSTANCES

No Commitment

No Flexibility

Can Get Terminated by AWS

Very difficult to Work With

Most Cost Efficient Option



RESERVED INSTANCES

1 or 3 Year Commitment

Low/Moderate Flexibility

Option for Upfront Payments

Difficult to Work With

Cost Efficient



SAVING PLANS

1 or 3 Year Commitment

Moderate/High Flexibility

Option for Upfront Payments

Easy to Work With

Cost Efficient

Improving the Cost of Storage

Don't forget EBS volumes! Identify those with low storage utilization, and reduce your costs by simply snapshotting them and then deleting them.

If you feel like storage is lacking governance and control, S3 analytics can give you a better view over the patterns in your storage, and can even recommend changes such as utilizing storage tiering, which can be used for both S3 and EFS.

For example, S3 Infrequently Accessed is great for applications and data that you rarely need, and can offer a much more cost-effective price per month over storage you need regular access to. You can leverage automation here too. For example, S3 Intelligent Tiering, automatically moves objects to the right storage tier based on how long it's been unused or accessed. It can also delete unused storage files using an S3 lifecycle rule, scheduling unused files to be deleted on a regular basis as defined by your policies and context. Don't forget EBS volumes! Identify those with low storage utilization, and reduce your costs by simply snapshotting them and then deleting them.

Ultimately, it's really important to align storage redundancy with your usage needs, using reporting and analytics to work with more than just a gut feeling about your environment.

Asking Yourself if Serverless is Right for You

Serverless reduces the cost of development, with no need to spin up costly servers, and dynamic auto-scaling is included as standard to meet real-time usage by running code-as-a-service.



Save **50%** on EC2 commitments and benefit from the full flexibility of the cloud, without any human effort.

[Learn More >](#)

Many organizations have found that utilizing serverless computing via [AWS Lambda](#) has helped them with cost optimization. From an architecture point of view, the compute, storage and network services can all be slashed, because infrastructure cost is based only on actual execution time.

Serverless reduces the cost of development, with no need to spin up costly servers, and dynamic auto-scaling is included as standard to meet real-time usage by running code-as-a-service. Serverless environments take on average 68% less time to provision than traditional cloud environments which AWS and Deloitte predict could equal hundreds of dollars per month per application.

Implementing Smart Automation to See Quick Time to Savings

Here's Where Zesty Adds the Flavor . . .

We've made it all sound easy, but we know that a lot of these processes take time, effort, and even holistic cultural change across the organization. Cost savings can't appear overnight. Except... when they can. Here's some examples of what Zesty can do - adding automation to the way you do business and solving these common optimization challenges faster than you can scroll up to remind yourself what that costing tool thingy was called. (It was AWS Cost Explorer.)

- ▶ **Get the Most out of EC2:** Most IT teams are stuck between a rock and a hard place, trying to keep costs to a minimum without stopping DevOps in their tracks when they've got their build on. Zesty uses AI to detect how much compute you need, and then buys and sells Reserved Instances according to exact needs. Save 50% on EC2 commitments and benefit from the full flexibility of the cloud, without any human effort.

The only tool of its kind in the market, it not only helps to ensure application stability, it also scales storage purchase to usage, enabling companies to cut their EBS costs by **70%**.

[Visit Zesty in AWS Marketplace >](#)

▶ **Optimize EBS volumes:** Even with the best will in the world, engineers can't stop over provisioning storage volumes. After all, they can't risk reaching the disk threshold and causing application failure. That's why Zesty offers an AI-powered auto scaling tool that shrinks and expands EBS volumes according to real-time needs, welcoming spikes in usage. The only tool of its kind in the market, it not only helps to ensure application stability, it also scales storage purchase to usage, enabling companies to cut their EBS costs by 70%. Let's just take that pressure off the engineer's to-do list altogether, eh?

▶ **Provides real-time insights into EBS and EC2 Usage:** All of these cost optimization tips start with visibility, and that's exactly what Zesty provides. While the AI functionality automates your cloud provisioning to reduce waste behind the scenes, you can check in and view exactly what you use and how. This allows you to focus on what humans do best; the big picture strategy for your long-term road map.

Already feeling like a cost optimization hero from all the knowledge we just dropped? Qualify for your official cloud costs mask and cape by [visiting Zesty in AWS Marketplace](#).

About Zesty

Zesty is the world's first AI-driven cloud management technology that auto-scales cloud resources to fit real-time application needs. As today's cloud environments become increasingly dynamic, Zesty automates cloud efficiency, improves DevOps productivity, and reduces cloud costs with zero human input. As a result, DevOps engineers no longer need to spend time on repetitive, and mundane infrastructure management tasks and can enjoy the cloud's flexibility and scalability without worrying about cost or maintenance concerns. Zesty was founded in 2019 in Tel Aviv and is used by leading organizations such as Armis, Gong, Yotpo, and others. For more information, visit [Zesty.co](#).

