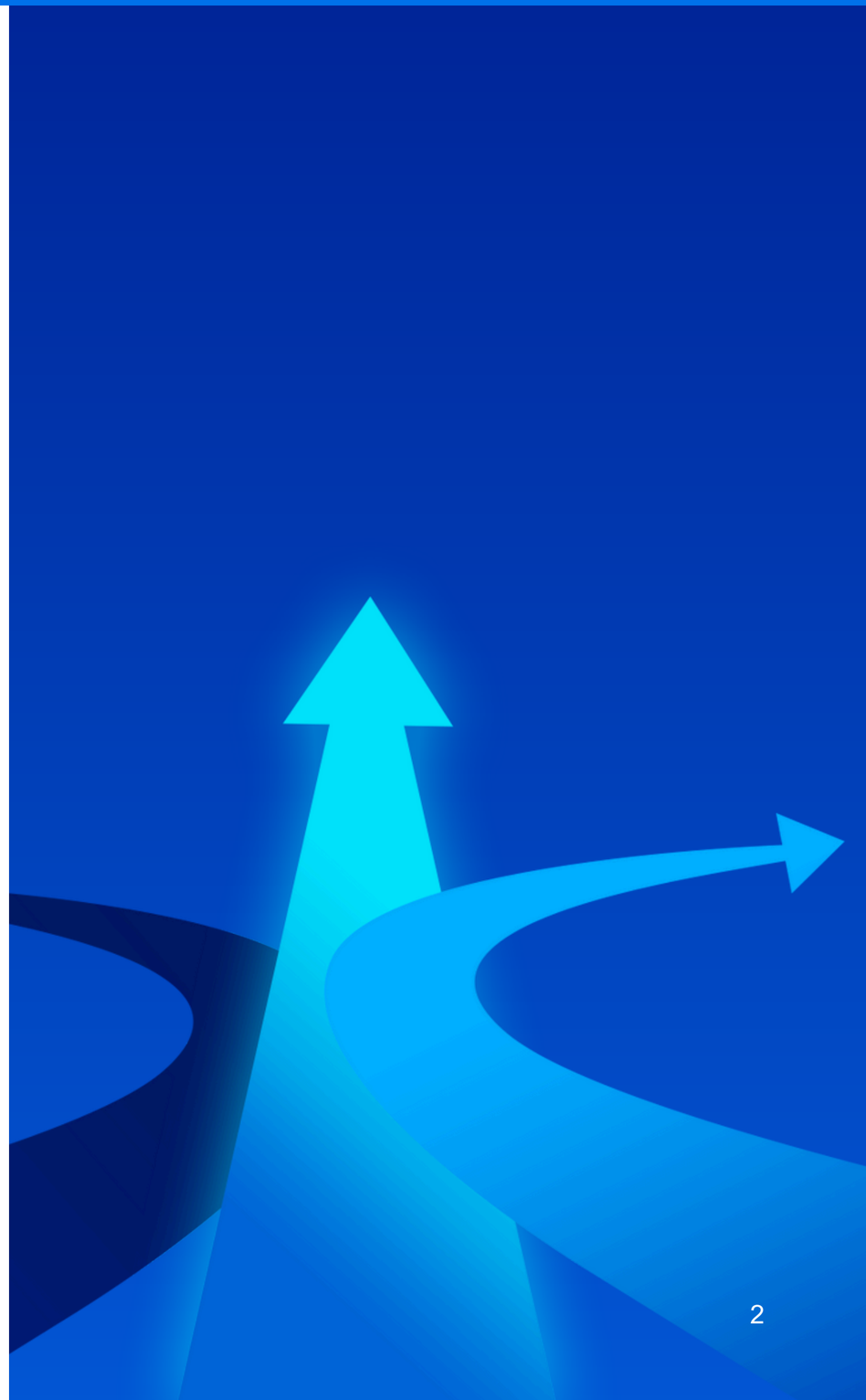
GUIDE

# Cloud Cost Optimization Buyer's Guide

# Table of Contents





# Introduction

With cloud spending making up a large and growing share of IT budgets, proper cloud management is crucial for planning and controlling costs. We've all heard the billing horror stories: the lone forgotten instance incurring thousands in cloud costs, or the startup taken down by unexpected overruns.

Everyone wants to save money in the cloud, and many platforms are being marketed to help you do just that. Yet, it can be challenging to decide on the best solution. The right platform can help you cut through complexity and spend less. On the other hand, selecting the wrong platform might lead to unpleasant billing surprises and missed savings.

The Cloud Cost Optimization Buyer's Guide was created to help business and finance leaders understand the features available in the current market and how to save with them. It covers:

- Top areas of cloud cost reduction in 2024
- Key considerations for tool selection
- Platform features and functionalities

Let's start by diving into the five key areas of cloud cost optimization.

## About this guide:

Many selection guides list popular solutions on the market with reviews and comparisons. We aim instead to help you understand the most important concepts and features, from right sizing to Spot instances. You'll learn the key criteria to evaluate during the buying process and their impact on you and your customers — so you can ask the right questions to assess a tool's suitability to your specific business needs.

# 01. Cloud Management: Cost Allocation, Visibility & Reporting

Optimizing cloud costs often starts with addressing the key issue of limited visibility into expenses. Without clear insights, it's hard to identify what and how much to cut. Key challenges typically include:

- The AWS bill is confusing, and it’s difficult to attribute shared costs
- It’s difficult to forecast spending and stay within the budget
- More visibility is needed to investigate unexpected charges or spikes in usage

The goal is to fully allocate, analyze, and report cloud costs so that you understand how resources are being used and by whom.

## Key Terms:

**Cost allocation:** the process of identifying and assigning the costs of shared cloud resources and services to different projects, applications, users or teams within an organization.

**Chargebacks:** a type of cost allocation in which the costs of cloud services and resources are billed to the business units or teams that use them.

**Tags:** simple metadata attached to resources in order to make them easily identifiable. If your AWS account has hundreds or thousands of resources, tags help you to categorize, filter and organize depending on different criteria such as project, environment or purpose.

**Showbacks:** a type of cost allocation in which the costs of cloud services are reported to the business units or teams that used them (but not billed).

# Features and Capabilities to look for

| Features | Capabilities |
|---|---|
| **Hourly Visibility** | With hourly granularity, you're able to see patterns and spikes in usage, allowing you to understand drivers of cost and plan commitments accordingly. A cloud cost visibility tool should allow you to break down cloud spending by service, resource, and other essential criteria. How easy is it to access and interpret this granular level of data? Does the tool provide real-time insights to help you react swiftly to unexpected cost spikes? |
| **Handling Tagging Misconfiguration** | Look for the ability to dynamic allocate rules by region, tags, operation, accounts, and usage types. Can you easily identify the percentage of untagged resources? Is there an option to automatically fix untagged or mistagged resources? |
| **Chargebacks & Showbacks** | Often, shared costs such as bandwidth or data transfer are difficult to allocate to specific projects or teams. How straightforward is it to assign costs to specific departments or projects with the platform? Look for features that facilitate accurate billing or cost reporting based on actual usage. |
| **Alerting & Notifications** | Determine the robustness of the tool's alerting and notification system. Can you set custom thresholds for spending alerts? Check for the ability to receive real-time notifications on cost overruns or unusual spending patterns. |
| **Customizable Reporting** | The ability to customize dashboards and reports is vital to meet your company's specific monitoring and reporting requirements. Evaluate how flexible and user-friendly these features are. Can you easily create reports to track key metrics for different stakeholders? |

# 02. Commitment Management: Reserved Instances & Savings Plans

In exchange for a commitment to spend a certain amount, AWS will offer a substantial discount on compute resources. However, it’s not always easy to determine the right type or amount of commitment to make to save the most. If you under-commit, you miss out on discounts. But if you over-commit, you’re paying for resources you don’t need. Let’s break down how to find a platform that can help you achieve the optimal balance.

## Understanding the 4 AWS pricing options

**On-Demand** is the most flexible model, but the most expensive. You simply tell AWS what kind of cloud resources you need, at the moment that you need them. Your goal should be to have 0% of your compute running on-demand.

**Reserved Instances (RI)** offer a substantial discount (up to 75% off On-Demand), but the least flexibility. Standard RIs require a commitment to a specific type of compute (instance family, region, operating system…).

**Savings Plans (SP)** offer discounts (up to 72% off On-Demand) in exchange for a commitment to using a certain amount of compute over 1 or 3 year periods. They offer additional flexibility compared to RIs, and are automatically applied by AWS to the spend that will result in the greestest discount.

**Spot instances** are spare AWS capacity that users can purchase at a heavy discount. The trick is that AWS may need the capacity back at any time — potentially disrupting workloads if not managed properly.

# Why is it challenging to optimize commitments?

Forecasting costs and planning commitments is an art. There are many possible choices and combinations — for example, there might be over 200,000 pricing options for one particular type of compute in a single region. And as business needs determine cloud usage, it’s necessary to continually rebalance and re-optimize. Determining the right commitments to make requires often analyzing the past 30-45 days of historical data, which takes time.

In 2024, AI solutions are available to continually monitor your usage and update your commitments on the fly — allowing you to avoid the time-consuming and costly process of doing it yourself. Here are the features to look for.

# Features and Capabilities to look for

| Features | Capabilities |
|---|---|
| **Organization-Wide Monitoring** | One benefit of AWS commitments is that they can be applied across accounts, teams, and departments. That's why organization-wide monitoring is crucial to maximizing cost savings. Does the tool offer real-time tracking of costs and usage? Assess the ease of accessing these insights and their ability to inform your purchasing decisions. |
| **Smart Purchase Recommendations** | Understand the tool's capability to offer tailored recommendations purchases of Reserved Instances (RI) and Savings Plans (SP). How effectively does it analyze usage patterns to suggest what and how much to buy? |
| **Commitment Guarantees and Buybacks** | The right platform can offer you more flexibility in adjusting to changing needs. Does the solution guarantee that it won't over-commit you to cloud resources? Does it offer buyback options for unused resources? |
| **Comprehensive Service Coverage** | Check the extent of the tool's coverage across the cloud services most important to your organization (i.e., the biggest line items on your AWS bill). Does it help you save on services like EC2, EKS, RDS, ASG, Data Transfer, Batch, and more? |

# 03. AWS Spot Instance Optimization and Automation

AWS Spot Instances are spare AWS (On-Demand)capacity that users can purchase at a heavy discount. Let’s discuss why your organization might need Spot, why it can be challenging to manage, and how tools can help.

## Spot FAQ

### Q1 Do I need to use Spot?

If you want to achieve the highest discounts on your compute, it is necessary to use Spot. Often, teams think they can cover 100% of their compute by buying Saving Plans or Reserved Instances. However, it’s typically only cost-effective to cover about half of On-Demand usage with commitments.

Why? SPs and RIs apply hourly, and operate on a “use it or lose it” basis. If you commit to $10 every hour but only consume $6 of services in a particular hour, you will lose the remaining $4. Because you can’t cover usage spikes with commitments, most companies are left with 40-50% of compute usage On-Demand, the highest AWS pricing tier. Spot is the only way to achieve discounts on this usage.

## Q2 Why can it be challenging to use Spot?

AWS gives you a discount on the instance, but not a guarantee that you'll be able to use it to the end of your compute need. The instance can be taken away at any time with a 2-minute warning, potentially disrupting critical workloads.

In addition, the Spot market is highly volatile. It is complex to manage fluctuating pricing, the types of instances available on the market, and the most cost-effective decision to make at any time, in view of existing commitments across your organization.

## Q3 How reliable are Spot instances?

Many organizations we've talked to believe Spot terminates at an alarmingly high rate (50% of the time!?) — but the reality is that less than 5% of Spot instances are terminated in any given month. Many strategies are available for using Spot reliably, including (1) designing application to be fault-tolerant (2) using multiple diverse instance types (3) prioritizing instance types that are historically less likely to be terminated.

Luckily, tools can reduce much of the complexity involved in using Spot. By leveraging Machine Learning to analyze the market, your usage, and other dynamic factors, they can put you on the most reliable and cost-effective Spot options automatically.

# Features and Capabilities to look for

| Features | Capabilities |
|---|---|
| **Advance termination notice** | Timely alerts for Spot instance terminations are essential. How far in advance does the tool notify you of impending terminations, and what mechanisms are in place to handle these terminations? |
| **Automatic management of instances** | By automating the lifecycle of Spot instances, a tool can help make it faster and easier to save with Spot. How efficient is the tool in launching, managing, and gracefully terminating Spot instances based on your workload requirements? |
| **Commitment awareness** | For the most cost savings, the tool must be aware of commitments across your whole organization. Look for the capability to balance between Reserved Instances (RI), Savings Plans (SP) and Spot instances to maximize cost savings automatically. How well can the tool forecast optimal commitment purchases based on your historical data and dynamic usage? |
| **Risk mitigation of interruptions** | Assess the tool's strategies for mitigating risks associated with using spot instances. How does it ensure continuity and stability of your workloads in the face of instance interruptions? Look for features that offer automatic failover, backup solutions, or seamless transition to on-demand instances to maintain operational integrity. Consider the tool's ability to predict and avoid potential disruptions based on market trends and data analysis. |
| **No vendor lock-in** | A tool that integrates with the AWS native services that you're using may be preferable to a proprietary scaling framework. Does the tool require you to onboard to their scaling framework or modify your current operations? Is it easy to change your mind and switch to a new tool if necessary? |

# Features and Capabilities to look for

| Features | Capabilities |
|---|---|
| **Right sizing recommendations** | Evaluate the tool's capability to analyze your current cloud infrastructure and provide actionable recommendations for right sizing. Ensure that the tool has a means of collecting your historical data on memory, utilization, network bandwidth, and volume size to provide trustworthy and reliable recommendations. |
| **Integration with your Infrastructure as Code (IaC)** | Assess the ease with which you can implement the tool’s right sizing recommendations. Does it integrate with your Infrastructure as Code (IaC) for one-click application of these suggestions, streamlining the optimization process? If not, applying recommendations may require significant time investment. |
| **Scheduling** | You can save on cloud costs by pausing resources that are idle or not actively needed at certain times. For example, you might want to pause UAT or Dev environments on nights or weekends. Explore the tool's scheduling capabilities — does it allow you to automate the start and stop of instances based on your workload schedules? |
| **Storage optimization & migration** | One key way to save is by automatically migrating from gp2 to gp3 storage volumes. However, the process can be time-consuming without a tool to identify eligible volumes for migration and automatically perform the migration. This feature can be a significant cost-saving measure due to the pricing and performance advantages of gp3 over gp2. |

# 04. Cost Optimization Automation: Right Sizing & Scheduling

As your usage patterns evolve, you'll find that some applications no longer need as many resources available to them as they once did. It's essential to continually right size these applications so that you're not overpaying for oversized or unnecessary instances. By analyzing historical usage and performance, you can identify and eliminate cloud waste and unnecessary spending.

## Common right sizing challenges:

**#1** Right sizing recommendations are frequently untrustworthy. Vast amounts of right sizing recommendations are available through tools such as AWS Trusted Advisor. Yet, it is often difficult to know which recommendations to accept. Reliable recommendations require granular historical data on memory, utilization, network bandwidth, volume size, and more. This data is difficult to collect and analyze, and there often isn't a single source for all of the metrics that you need.

**#2** Acting on recommendations is time-consuming when done manually. It is often necessary to investigate, create tickets, and implement recommendations. And because right sizing is an ongoing process, it can consume considerable engineering time and effort.

# 05. MAP (Migration Acceleration Program) Tagging & Tracking

Migration Acceleration Program (MAP) offers incentives to help AWS customers accelerate their migration to AWS from on-premises workloads to AWS.

## How does AWS MAP work?

To identify the migrated workloads and track their spending on AWS, a tagging mechanism is employed. It allows you to effectively track the spend associated with the migrated workloads. AWS then calculates financial credits based on this spend and applies those credits to your overall AWS bill to support your cost of migration. The more workloads you migrate, the more incentives you receive.

# Features and Capabilities to look for

| Features | Capabilities |
|---|---|
| **Automatic Tagging of MAP Resources** | It’s very important to properly tag resources, because it gives AWS visibility into your migration process. To maximize MAP funding, make sure all of your resources are tagged properly. Ensure the tool can automatically and accurately tag resources associated with MAP for proper tracking and management of your migration. |
| **Automatic Correction of Untagged and Mislabeled Resources** | Every day that you have resources that are not tagged properly, you lose money. Evaluate the tool’s proficiency in identifying and correcting untagged or mislabeled resources. How quickly and accurately can it detect such anomalies, and what automated measures does it implement to rectify these issues? |
| **Automatic MAP Tracking** | Assess the tool’s capacity to track and help you visualize your migration process. Does it provide a comprehensive dashboard with a unified view of your migration progress and the credits you have received? Consider the level of detail, real-time updates, and usability of the dashboard. |

# Cloud Cost Optimization Tool Selection Checklist

## Cloud Management

- [ ] Hourly Visibility
- [ ] Robust Tagging Features
- [ ] Chargebacks & Showbacks
- [ ] Alerting & Notifications
- [ ] Customizable Reporting

## Commitment Management

- [ ] Organization-Wide Monitoring
- [ ] Purchase Recommendations
- [ ] Guarantees and Buybacks
- [ ] Comprehensive Coverage
- [ ] EKS

## Resource Optimization

- [ ] Right Sizing Recommendations
- [ ] Integration with your Infrastructure as Code (IaC)
- [ ] Scheduling
- [ ] Storage Optimization & Migration

## Spot Automation

- [ ] Advance Termination Notice
- [ ] Instance Management
- [ ] Commitment Awareness
- [ ] Risk Mitigation of Interruptions
- [ ] No Vendor Lock-In

## Comprehensive Coverage

- [ ] EKS
- [ ] EC2
- [ ] RDS
- [ ] Auto Scaling Groups (ASG)
- [ ] Batch

## MAP Management

- [ ] Automatic Tagging
- [ ] Mislabel Fixing
- [ ] MAP Tracking

# Comparison of Cloud Optimization Platforms

Compare the key features offered by popular cloud optimization solutions in this summary table.

| Features | nOps | AWS CloudWatch | Apptio Cloudability | Spot by NetApp | CloudHealth | Zesty | CloudZero | ProsperOps | Cast.ai |
|---|---|---|---|---|---|---|---|---|---|
| VISIBILITY | | | | | | | | | |
| Cost Allocation | ● | ◑ | ● | ◑ | ● | ○ | ● | ○ | ◑ |
| Reports & Dashboards | ● | ◑ | ● | ◑ | ● | ○ | ● | ○ | ◑ |
| Budgeting | ● | ◑ | ● | ◑ | ● | ○ | ● | ○ | ○ |
| COMMITMENT MANAGEMENT | | | | | | | | | |
| Management of Reserved Instances & Savings Plan | ● | ○ | ◕ | ◔ | ◑ | ● | ○ | ● | ○ |
| 100% Utilization Guarantee | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| SPOT MANAGEMENT | | | | | | | | | |
| Real-time workload Management | ● | ○ | ○ | ● | ○ | ○ | ○ | ○ | ◕ |
| Spot market awareness | ● | ○ | ○ | ● | ○ | ◑ | ○ | ○ | ◕ |
| Commitment awareness | ● | ○ | ○ | ◕ | ○ | ○ | ○ | ○ | ○ |
| Support for EKS/ECS/ASG | ● | ○ | ○ | ◕ | ○ | | | | ◔ |
| CLOUD OPTIMIZATION | | | | | | | | | |
| Resource Rightsizing | ● | ◔ | ◑ | ◕ | ◕ | ◕ | ◔ | ○ | ◑ |
| Storage Optimization | ● | ○ | ◑ | ○ | ◑ | ◑ | ○ | ○ | ◔ |
| Resource Scheduling | ● | ○ | ◕ | ○ | ○ | ◑ | ○ | ○ | ○ |
| ADVANCED EKS FEATURES | | | | | | | | | |
| Support for Native Autoscalers (CA/ Karpenter) | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| Container and cluster efficiency dashboards | ● | ○ | ○ | ◕ | ◑ | ○ | ○ | ○ | ● |
| Node & Container Efficiency insights & automation | ● | ○ | ○ | ◕ | ◑ | ○ | ○ | ○ | ● |

# About nOps

# About nOps

nOps delivers you comprehensive AWS cost savings, while freeing up your time to focus on building and innovating. A few of the benefits include:

### Cost Savings.

Our AI has real-time awareness of the Spot market, your existing commitments, and your changing usage. It automatically schedules the optimal instance types for your workloads, balancing between Compute SP and Spot to save you the most money.

### Complete solution.

We offer all of the features in the checklist above — no need to sign up for additional tools or subscriptions.

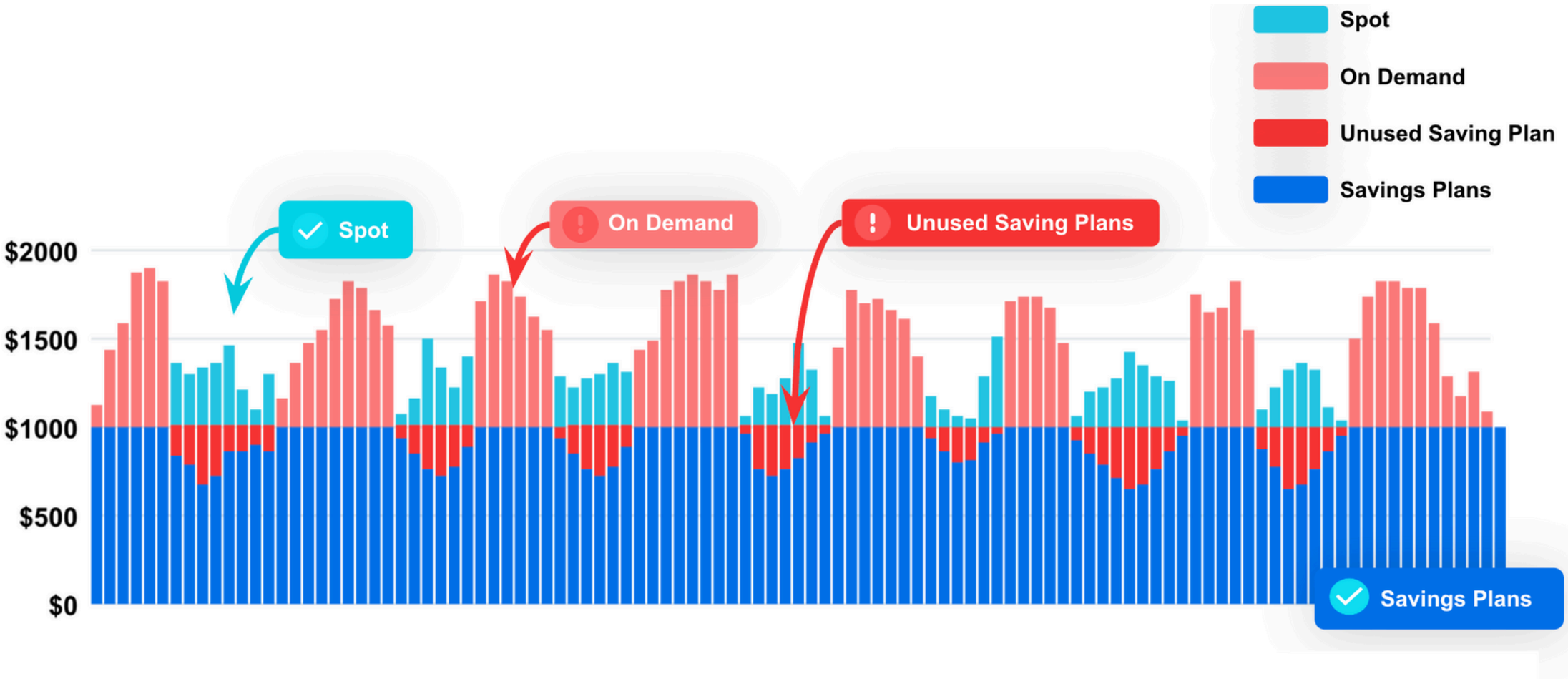
### No upfront costs.

The nOps platform is free to sign up and use. We only charge you a percentage of what we save you when you act on our recommendations or enable the optimization engine.

nOps is entrusted with over a billion dollars of AWS spend, and the platform was recently ranked #1 in G2's cloud cost management category.

Join our customers saving up to 50% by **booking a demo** today!

**Before:** **Unused commitments, inefficient Spot usage, and high On-Demand prices drive up AWS bills.**



**After:** **Copilot automatically balances your workload across Spot and SPs, ensuring all of your usage is discounted to maximize savings automatically.**