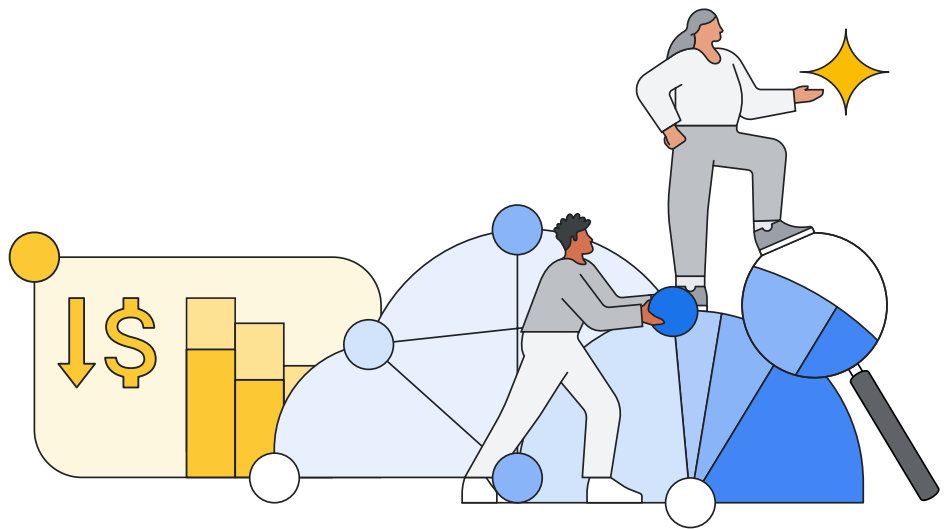Google Cloud

Achieving cloud financial resilience with Cloud FinOps

From quick fixes to reduce your cloud spending to implementing effective cost governance, Google Clouds' insights can help you develop your optimized, long-term cloud cost strategy.

Authors: Sheri Cunningham, Sam Moss, Pathik Sharma, Eric Lam, Tom Nikl

Google Cloud

Google Cloud is here to help you achieve financial resilience– now, and in the future.

We all know that the greater economic climate can change quickly: Unpredictable market conditions. Rising inflation. Increasing energy, import, and export costs. Supply chain disruptions. It's no wonder businesses worldwide are focused on optimizing costs and restructuring for greater financial resilience. Google Cloud is uniquely differentiated to help solve these challenges, and we're here to help.
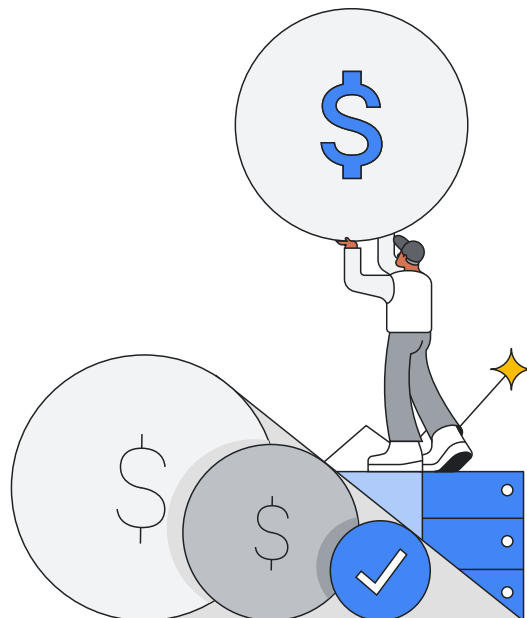
Don't let your investments get lost in the cloud.

It is commonly known that the cloud enables businesses to reduce their IT capital expenses. But, cloud done right can also help you reduce your operational expenses, while helping you continue to innovate and build new capabilities as well. In fact, by 2030 it's forecasted that Fortune 500 companies will realize over $1 trillion in value from cloud.

However, respondents to the Flexera™ 2022: State of the Cloud survey estimate **their wasted cloud spend is still above 28%**[2], and a contributing factor is utilizing on-premises computing calculations and procedures in the cloud. Bottom line: we're here to help you drive maximum business value by finding creative ways to reduce (or eliminate) wasted cloud spend so that your OpEx is optimized as well as your CapEx.

Did you know? 🔍

In fact, by 2030 it's forecasted that Fortune 500 companies will realize over **$1 trillion** in value from cloud.[1]

[1] McKinsey: Cloud's trillion-dollar prize is up for grabs, February 2021. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/clouds-trillion-dollar-prize-is-up-for-grabs

[2] FLEXERA™ 2022: State of the Cloud Report. https://resources.flexera.com/web/pdf/Flexera-State-of-the-Cloud-Report-2022.pdf?elqTrackId=f3bb660986704d2980404386aa003141&elqaid=6925&elqat=2

Google Cloud

Your FinOps journey starts here.

First, what is Cloud FinOps? Cloud FinOps is an operational framework and cultural shift that brings technology, finance, and business together to drive financial accountability and accelerate business value realization through cloud transformation. FinOps is a practice whose goal is to gain visibility into your cloud costs, maximize your cloud investments, and develop cost-conscious practices across teams. FinOps isn't just another checklist of procedures – it's a way of reimagining your organization as a cloud-first entity. If you're interested, here's a deeper dive on FinOps.

In this guide, we'll show you how to start your organization's FinOps journey. The destination?
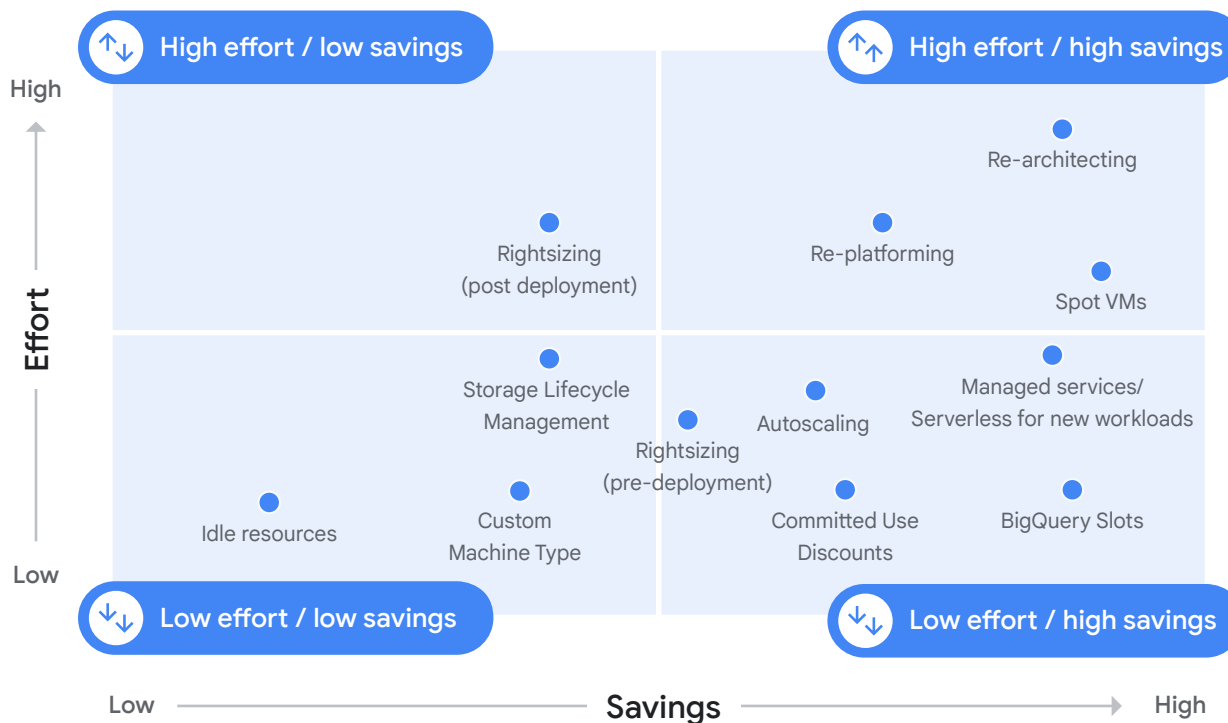
A cloud-powered machine where teams share resources and workloads – resulting in lower resource costs, reduced labor needs, and improved data and insights that drive business innovation. But, we know that journeys often happen in phases. So we'll start with five quick ways you can start cost optimizing today. And then we'll cover four longer-term ways to continue your FinOps journey.

Take advantage of these five quick wins

Prioritizing low-effort changes helps your organization make the greatest impact, quickly. As you continue on your FinOps journey, you can implement higher-effort changes for an even larger impact. Let's look at a few key cloud cost optimization approaches that you can implement right away.

Cloud cost optimization matrix: effort vs savings

High effort / low savings

High effort / high savings

High

Effort

Re-architecting

Rightsizing (post deployment)

Re-platforming

Spot VMs

Storage Lifecycle Management

Autoscaling

Managed services/ Serverless for new workloads

Rightsizing (pre-deployment)

Idle resources

Custom Machine Type

Committed Use Discounts

BigQuery Slots

Low

Low effort / low savings

Low effort / high savings

Low   Savings   High

Google Cloud

1 Remove idle resources

Idle resources can occur for a number of reasons. Maybe you started a test environment that is no longer used, or maybe an environment just isn't getting the usage it once was. Either way, idle resources are a great place to start trimming excess cloud spend if you can find them. That's why Google Cloud gives you proactive alerts when things are idle, including the following features included in our Active Assist portfolio:

Unattended project recommender is a tool that analyzes usage activity and identifies any projects which have had low usage for the last 30 days – a sign a project may have been abandoned or forgotten. These projects may still contain billable resources, and even if not, you may want to shut them down to eliminate the security and cost risk.

Idle VM recommendations help identify VMs, persistent disks, IP addresses, and custom disk images that have not been used over the past 14 days. Since these recommendations appear in the Console UI at an individual project level, you may find it easiest to export these recommendations to BigQuery for a single view of the data.

Idle GKE Cluster recommendations help you identify when GKE clusters are unused.

PayPal is a strong example of achieving both short term and more strategic cost optimization results.

Through a multi-phased sprint to identify low effort opportunities, PayPal collaborated with Google and Deloitte on identifying idle resources for unattached disks, quarantined VMs, and aged snapshots. Labeling resources and taking comprehensive steps to ensure they could be deleted.

PayPal achieved

↓ **5% reduction**

in annual cost avoidance after only a few months of analysis and execution.

2   Rightsize your cloud resources

One of the best ways to optimize cloud spend is to make sure you're only paying for the resources that you're actually using. Seems simple enough, right? But it can be surprisingly challenging if you're not using the right tools, like:

- Compute Engine's rightsizing recommendations to optimize the resource utilization of your VM instances. The recommender also suggests custom machine types when appropriate – helping you save money by provisioning a custom machine with the exact number of cores and memory required.

- The Cloud SQL overprovisioned instance recommender helps you detect instances that are unnecessarily large for a given workload

- Google Cloud Storage Autoclass is a managed service that automatically moves data that is not frequently accessed to colder storage classes and delivers immediate cost savings. Autoclass also helps reduce the cost of ongoing accesses by removing early delete and retrieval charges for your data in colder classes.

"

Not only would it cost valuable engineering resources to build cost-optimization ourselves, but it would open us up to potentially costly mistakes in which we incur retrieval charges for prematurely archived data. Autoclass helps us reduce storage costs and achieve price predictability in a simple and automated way."

Ian Matthews,
CEO, Redivis

Google Cloud

Google offers multiple types of discounts, including committed use discounts (CUDs). CUDs are like buying in bulk – spend more, save more. You can apply them to multiple parts of your application infrastructure, including virtual machines, databases, and more. There are 2 types of CUDs:
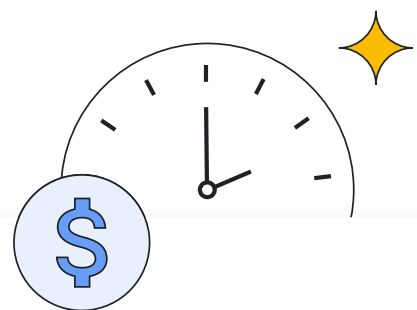
$ **Spend-based:**

Commit to a certain level of spend per product for either 1 or 3 years on a Google Cloud billing account, and receive up to 40% off. Analyze your product usage to identify areas with a minimum predictable spend – using billing console reports or the billing export.
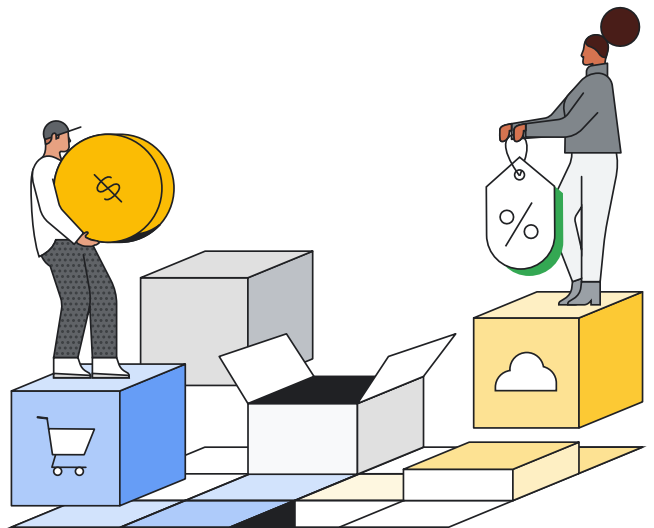
⚙ **Resource-based:**

Commit to a certain level of resource usage, and get an even deeper discount (up to 57% for standard CUDs). Only certain products support CUDs – you can see what's supported here. To identify which are right for your organization, use the recommender.

CUDs can be quickly implemented because they don't require changes in resource usage. These discounts can be activated in the Console UI and can stack on top of any Enterprise Commit Deals already negotiated with Google.

Etsy took advantage of Google Cloud's CUDs – lowering its compute cost by **42%** alongside other optimizations.

Google Cloud

4 **Move to Managed Services**

Google Cloud offers many managed service versions of the workload you're running, which often result in a much stronger ROI than running them on-prem. Just a few examples include GKE for containers, Cloud SQL and Memorystore for databases, and BiqQuery for data warehouses.

Diving into our BigQuery example, BigQuery slots are units used to execute workloads, and there are two different pricing models available: on-demand and flat-rate. Your organization can save significant costs by selecting the pricing model that's right for your needs. Because on-demand bills by queries scanned, it's the most cost-efficient model if your usage is unpredictable and often low. If your usage is consistent and significant, the flat-rate model is best – with fees based on your calculated needs and discounts depending on the length of commitment.

You can also leverage Active Assist's BigQuery slot recommendations, which help to further optimize cost when using BigQuery. This feature helps give you automatic answers to questions like "How many slots do I need?" based on your BigQuery usage history, spend, and other signals. With actionable and automatic insights and recommendations, you can easily understand the cost benefits and performance tradeoffs of switching to the longer-term reservations for a given project or organization.

Did you know? 🔍

General Mills took control of their FinOps fate by strengthening their cost governance. After discovering that a Google Cloud project was accruing excess charges, the engineering team identified a lack of partitioning and clustering of their BigQuery datasets, rectified it quickly, and saved over $80,000.

5 **Use billing reports**

One question you may grapple with is "how do your teams know how much they're spending if they can't see the numbers?" In Google Cloud, you have access to Console Billing Reports that provide visibility, can be used to analyze spend, and **help you answer questions such as:**

- How is my current month's Google Cloud spending trending?
- Which Google Cloud project had the highest cost last month?
- What are my forecasted future costs based on historical trends?

Getting started is easy; simply enable these reports as appropriate and grant the Billing Account Viewer role to the respective individuals. For even greater ease of use, you can also enable the BigQuery Billing export so billing data can be exported to a BigQuery dataset – providing more flexibility when analyzing or reporting on spend in the future.

Google Cloud

Continuing with a strategic approach to cost optimization: 4 key practices

Starting with our five quick wins above is a great place to start, but as we mentioned, there is much more that can be done. To continue your cloud FinOps journey, we've put together four longer-term ways that you can fine tune strategically as you continue to grow in the cloud. These curated leading practices are the result of Google Cloud helping many customers optimize their cloud cost management practices over the years.

1 Assess and organize

As an organization grows, it can become more difficult to track cloud resources – leading to cloud sprawl and waste. To keep your cloud resources efficient, design a robust resource hierarchy, implement a cost allocation strategy, and apply automations as guardrails for cost control strategies.

Resource Hierarchy

Establishing a reliable resource hierarchy – where departments, teams, and project folders are self-contained and organized – is crucial for cost governance. Without a strong hierarchy, attributing spend and business value to different workloads is more challenging as all information needs to be maintained in metadata (labels or tags) attached to resources in projects.

Cost Allocation

Once your resources are organized using the hierarchy, you can start applying metadata to projects and resources so that you know more precisely what is running within a given folder. You don't want the equivalent of a desktop with 'new folder' and 'new new folder'!

Tagging (or labeling) resources allows you to easily flag resources by department, owner, environment, and more.

Cloud Tagging Framework

Security tags
- Name
- Compliance/ criticality
- Classification
- Encryption

Tags for automation
- Enddate
- Criticality
- Schedule-off
- Schedule-on

Business tags
- Application
- Cost center
- Criticality
- Project-ID

Technical tags
- Owner
- Enddate
- Environment
- Name
- Function
- Release

When resources are tagged in this way, you're only a quick billing report away from identifying the operational costs for a certain product or even department – providing cost transparency and giving you the ability to make informed decisions on cloud investments.

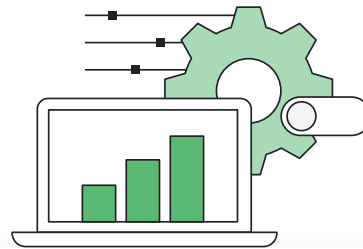**There are two options for tagging resources in Google Cloud:**

Labels in Google Cloud:

Labels are simple key-value pairs which you can apply to many resources in Google Cloud. They appear in the billing export and have been historically how to manage cost allocation. Labels can be useful but they also lack certain features which are beneficial for cost control, for example governance, inheritance and binding.

Google Cloud Tags:

Tags are a newer feature in Google Cloud and are more feature rich than labels. They make allocating cloud costs easier through improved governance tools, binding and inheritance. Because tags are new, we recommend using tags over labels if you are building your framework from scratch, but maintaining labels for existing projects and phasing out over time. For more information on which products Tags currently support visibility of in the billing export, please refer to our public documentation and whitepaper on application of Tags.

How can your organization start automating?

Identify the data sources – cloud logging, monitoring, audit events, etc.

Set triggers based on a predefined schedule or an event-based transaction.

Use conditions to filter transactions.

Execute the action – a recommendation, notification, or a task.

One example of an effective cost control strategy is setting up automated budgets and alerts to track and control costs – enabling your organization to track actual Google Cloud spend against planned spend. By configuring budget alert threshold rules and email notifications, your organization can stay informed about spend versus budget, or automate cost control responses to selectively control usage for the production vs non-production workloads.

Cost control strategies

We've all had the experience of doing a vast clean-up task only to see it return to a mess in a short amount of time. The key to maintaining your resource hierarchy and labeling on a long-term basis – and ensuring cost control – is to implement automation and governance policies at scale.

The exact proactive cost controls that are best for your organization depend on a number of personalized factors. We can be your partner in this process, helping your organization create a guidebook of tailored cost control policies. From there, we design an end-to-end workflow to automate based on triggers – so that your organization is proactively protected against cost sprawl.

2   Achieve unified visibility

If you can't measure it, you can't improve it. Making the right decisions to optimize or invest in the cloud requires complete visibility into your cost and usage.

How can I offer self-service cost reporting at all levels of my organization?

The right reporting will reveal cost trends and support use cases such as persona-based reporting, resource utilization, cost insights, and granular forecasting – which are fundamental to almost all aspects of cloud FinOps. Google cloud native tooling (cloud billing console reports, pricing calculator, and more), third-party SaaS solutions, and in-house solutions are all methods of reporting and visualization.

The average organization uses 3.7 tools for its reporting, forecasting, and optimizing needs, according to the 2023 State of FinOps survey by FinOps Foundation. Every time data is passed between non-natively-compatible products, there is a chance for missed insights, lack of integration, and data loss. Having a single, integrated tool eliminates these compatibility concerns and improves business value.

Google's product Looker Pro provides a single source of visibility, including cloud cost insights templates which can accelerate your adoption of in-house solutions. These solutions harmonize data across multi-cloud architecture for a unified view of cloud spend.
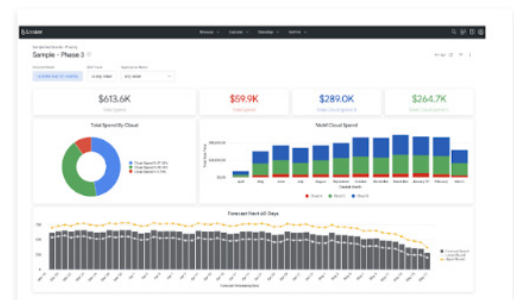
Exporting to BigQuery

Cost and usage data

- Billing export (S/D)
- Pricing export
- Asset inventory
- Recommendations export
- Utilization export
- Logging export
- Other cutom datasets (Marketing dataset, CMBD, etc.)

BigQuery → Looker / Looker Studio / Bi Platform

Billing data can be exported to Looker / Data Studio and other analytics tools to craft custom dashboards that dive deeper into your cloud usage and costs.

Google tools provide integrated visibility, reporting, and recommendations

Google Cloud

3 Gather cost insights

With clarity on cloud costs across teams and cost-centers, the next step is to leverage intelligence to analyze usage patterns and gain insights to help drive operational savings.

★ **Google's AI/ML features do the heavy lifting for you.**

One example, which we've mentioned in some of the quick wins above, is Active Assist, a portfolio of intelligent tools that helps you optimize your cloud operations with recommendations to reduce costs and increase performance.
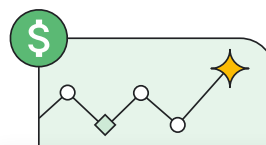
Did you know? 🔍

Google Cloud recently commissioned Forrester Research for a study which found: "Using Active Assist, organizations right-sized virtual machines (VMs) and removed idle VMs to optimize overall cloud consumption costs on Google Cloud." Additional, "for impacted workloads, Active Assist reduced overall consumption by 40%, resulting in a three year present value cost savings of $2.5M."
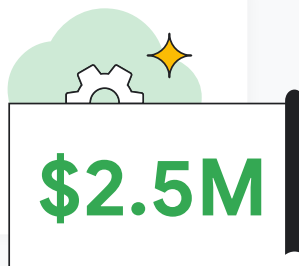
ROI
781%

Benefits PV
$2.6M

NPV
$2.3M

Benefits (Three-Year)
Optimized cloud consumption

$2.5M

4   Improve cost awareness across teams

Once your organization finishes labeling resources, increasing visibility, and improving insights – solidify your gains by creating a playbook to document best practices and develop cost-conscious skills across teams.

Ready to build your own playbook?

Learn more about Google Cloud Architecture Framework for cost optimization to get started. And here are some helpful resources on optimizing cost:

- Compute, containers, and serverless
- Storage
- Databases and smart analytics
- Networking
- Cloud operations

Cloud cost optimization is an ongoing consideration for your organization. These recommendations are an opportunity to engage closely with your cloud architecture and enable cross-functional collaboration that optimizes costs across product and engineering teams.

Google Cloud Cost Optimization Approach

↯↑ Discovery	⌁ Analysis	✄ Document & Review	⋀ Plan & Design	✔= Implement
Understanding Goals	Top Service / Project / Apps	Identify Opportunities	Prioritize Recommendations	Track Implementation
End to End Business Flow	Performance Metrics	Potential Savings ($)	Implementation Playbook	Define Savings KPIs
Architecture Review	Recommendation Signals	Estimated Level of Effort	Scripts / Code Templates	Reinvest Savings Realized ($)

⧑ Governance	Architecture Hub	Guardrails & Cost Control Policies	Cost Optimization Insights & Dashboards	Cost Optimization Champions

Getting from A to Business value

While it's critical to constantly optimize and drive billing trends down, there may be additional factors – like increased customer activity – that can offset savings. Even if the cost curve remains flat, value generated from the cloud cost must still be captured.

In a transformation with Google, OpenX was not only able to identify their per-unit cost in the cloud, but also to showcase a 60% reduction. This example shows that through careful planning and measurement, organizations can achieve and prove cloud-enabled business value. **Achieving a total cost of ownership per-unit metric (such as with OpenX as per advertising transaction), is an ideal way to define cost efficiency within an organization.** Granted, achieving an accurate unit metric view may not always be possible due to cloud maturity and the ability to capture spend across public and private clouds, as well as on-premises software.

FinOps practitioners often receive feedback from the lines of business that it is unclear how cost saving measures contribute to the company mission or success. In fact, daily optimizations are not commonplace in sunk cost capital purchases for on-premises technology. Shifting the mindset to align measurement as well as ties of cloud costs and value to the company mission is one way for an organization at any maturity level to begin to develop a culture of understanding and clarity on cloud benefits.

Begin with the organization's core objectives for the next 1-3 years. **For example, a clothing retail company** could have 3 strategic objectives:

- Transform into a data company that sells clothing

- Delight customers through custom in-store experiences

- Drive sustainable operations and brand efficiency

These objectives should be incorporated into cascading team and individual employee objectives. A key way to align cost savings to tangible business value is to tie efforts with **incentivized strategies**.

— Unit cost

▬ Total Cost

60% reduction
in per unit costs

Unit cost reporting

1 Transform into a data company that sells clothing

This strategy is aimed at understanding each customer, their journey, and their buying decisions. IT infrastructure must support this effort to store and harmonize with cross device and channel data. Individuals within the company are already incentivized to achieve this company goal successfully. How can one prove that they have effectively transformed and adopted a strong data estate? Take the overall cost of housing and storing the data – from storage, to analytics, and visualization – to develop a success metric of total technology cost of data and analytics.

Cost optimization levers discussed above, such as BigQuery Slots, can drive down those costs. Thus, the cost optimizations are related back to the overall transformation objective. One can, for example, achieve a percentage of savings within that transformation and exhibit this number as the value derived.

Taking this example a step further to not just direct total costs, value can be extracted from other metrics on the business value spectrum. See Maximizing Business Value with Cloud FinOps for a more in depth discussion of this framework.

Business value metrics that quantify ROI on cloud transformation

Cost Efficiency
Measure cost efficiency through infrastructure, savings, migration and support costs

Infrastructure

Support

Implementation

Resiliency
Enhance operational resiliency with improvement in service quality and security risk posture.

Service Quality

Secure Posture

Operational Stability

Velocity
Increase time to market by accelerating fluidity in product and service delivery

Developer Productivity

Release Frequency

Business Agility

Innovation
Enable a culture of rapid experimentation to drive innovation and cloud transformation

Return on innovation

Employee Experience

Customer Satisfaction

Sustainability
Embed true environmental and social sustainability across the organization

Carbon Footprint

Power Usage Effectiveness

Circular Economy

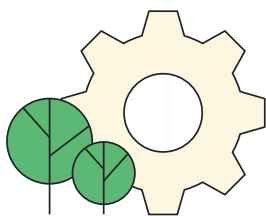Now companies can turn to advancing their other corporate objectives.

Google Cloud

2 **Delight customers through custom in-store experiences**

This can be tied to other business value metrics like return on innovation under the Innovation pillar. Within this example, the organization's cloud engineering team designs a pipeline to map in-store experiences to sales advertised via their mobile application. The shopper is delighted with showing their mobile app coupon to earn 20% off in-store. Collecting point-of-sale data and tying it to customers who have viewed sale items over the past 7 days will produce a benchmark of app users versus non-app users. These calculations will highlight the uplift from tying those experiences to an innovative data pipeline. This example can be used to advocate around the organization for not only lower spend derived from achieving cost efficiencies, but also for showcasing how unused funds can be allocated to higher business value driving activities.

Organizations can operate and design cloud implementations and methodologies to exhibit value in any of the business value metrics.

3 **Drive sustainable operations and brand efficiency**

Cost efficiency relates strongly to sustainability objectives. If a company has tighter governance and protocols to reduce waste, this results in a direct reduction in their carbon footprint and share of power utilized in the cloud. Organizations driving short or long-term FinOps initiatives can help to justify and garner clarity and support for their actions by aligning to their company's overall priorities.

Cost Efficiency Scorecard

Company objective	Sample metric
Transform into a data company that sells clothing	Total Cost of Data and Analytics Technology
	Total Cost of Ownership for Data and Analytic
Delight customers through bespoke, in-store experiences	In-store conversion rate for retailer mobile app shopper vs in-store conversion rate for unknown shopper
Drive sustainable operations and brand efficiency	Indirect greenhouse gas emissions from cloud footprint

Now companies can turn to advancing their other corporate objectives.
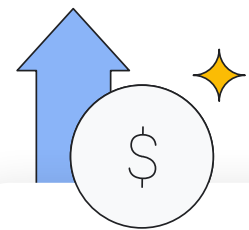
Google Cloud

Wrapping up

Instilling a continuous cloud cost optimization discipline is one of the key contributors to your organization's cloud transformation success. We've seen time and again that realizing the business values of the cloud **begins with a cloud-first culture.**

By transforming your company into a cloud-first organization, you can optimize resource costs and leverage the data insights to drive business innovations. For many organizations, this cultural shift comes with growing pains, but the destination is clear: visibility into cloud cost, maximizing your cloud investments, and developing cost-conscious practices across teams.

We've shown you the key principles that will guide your FinOps strategy, as well as some quick wins that will help you optimize cloud costs and manage cloud sprawl immediately. All you need to do now is get started. Interested in partnering with Google to develop your cost optimization strategy, implementation, and execution? Reach out to your account team to engage our FinOps Consulting practice.

Or, if you're looking for some more FinOps reading, take a look at these resources:

See Maximizing Business Value with Cloud FinOps for a deeper dive.

Interactive Tools:

Cost estimation calculator

Cost management tools

Active Assist Console

Training videos:

Understand your Google Cloud cost

Optimizing Google Cloud Cost | GKE

Optimizing object storage costs in Google Cloud: location and classes

Simplify and automate cost optimization with Autoclass for Cloud Storage

Customer videos:

Google Cloud Next 2020: OpenX Case Study

Google Cloud Next 2022: General Mills Case Study

Google Cloud IT Heroes 2023: Optimize for Shrinking Budgets

Whitepapers:

Framing up FinOps:
How to optimize your cloud costs on Google Cloud

Maximize Cost Efficiency with CUD Optimization Strategy

Principles of cloud cost optimization

Unlocking the value of cloud FinOps with new operating model

Driving Cloud FinOps at scale with Google Cloud Tagging

Getting started with FinOps on Google Cloud

Maximize your business value with Cloud FinOps

The business value of Google Cloud Storage, IDC

Google Cloud Architecture Framework: Cost optimization