The cloud economics playbook

A 3-6-9-3 framework to demystify cloud cost optimization

/thoughtworks

Strategy. Design. Engineering.

Jan 2025

The inevitable rise of cloud infrastructure

Cloud services are rapidly becoming indispensable for enterprises, offering significant advantages over traditional on-premises IT infrastructure.

- **Scalability and agility:** Cloud services provide on-demand resources that can be easily scaled up or down based on business needs.
- **Faster experimentation and innovation:** Cloud services provide access to cutting-edge technologies and tools that can help enterprises accelerate experimentation, leading to innovation and more efficient development of new products and services.
- **Reduced operational overhead:** By shifting IT infrastructure management to the cloud, enterprises can free up valuable IT resources with reduced operations overhead to focus on core business activities and strategic initiatives.

While some large enterprises may still maintain some on-premises infrastructure for specific needs, the overall trend is towards a hybrid model that leverages the benefits of both cloud and on-premises solutions. And with this increased cloud adoption comes a crucial factor: **cloud cost management**. Here's why managing cloud costs is essential for enterprises in today's cloud-centric world:

- **Underutilized and wasted resources:** The on-demand nature of cloud can be a double-edged sword. While it allows for easy scaling, it's also easy to overprovision resources that remain idle or leave resources that are unused.
- **Hidden costs:** Cloud services offer a vast array of features and options, making it easy to unknowingly incur unnecessary charges.

- **Shadow IT:** Experimentation and innovation are inherent aspects of a thriving cloud infrastructure. However, when developers spin up cloud resources outside of established guidelines to explore new services, it can lead to a lack of cost visibility and control — a practice often referred to as Shadow IT.
- **Unpredictable budgets:** Uncontrolled cloud spending can wreak havoc on financial planning. Without proactive cost management, enterprises face surprise spikes in bills and struggle to forecast future expenses accurately.

As cloud adoption continues to surge, managing cloud costs becomes the cornerstone of a successful and sustainable cloud strategy for large enterprises. With effective cloud cost controls, large enterprises can leverage the benefits of cloud computing without getting bogged down by financial mismanagement.

In this Playbook, let’s examine the 3-6-9-3 strategy of optimal cloud economics that helps with managing cloud cost and operations effectively.

Understand	**Avoid**
3	**6**
types of services	common pitfalls
Apply	**Uphold**
9	**3**
optimization strategies	pillars of governance

3

Three types of cloud services

Cloud services can be broadly categorized into three types, offering flexibility in deploying workloads. The best choice for your organization depends on many parameters, from application specific needs, in-house skills in managing operations, usage patterns and cost structure.

Serverless services, such as AWS Lambda, EMR Serverless, DynamoDB, SQS and SNS, AWS Fargate, and AWS ECS, are SaaS offerings from cloud providers. They eliminate operational overhead, including provisioning, configuration, maintenance and capacity planning. These services offer fast automatic scaling with a pay-per-use pricing model, which can become costly with higher usage.

Pro tip: Use a serverless service such as AWS Lambda for short duration tasks that run infrequently, such as daily policy check automation etc.

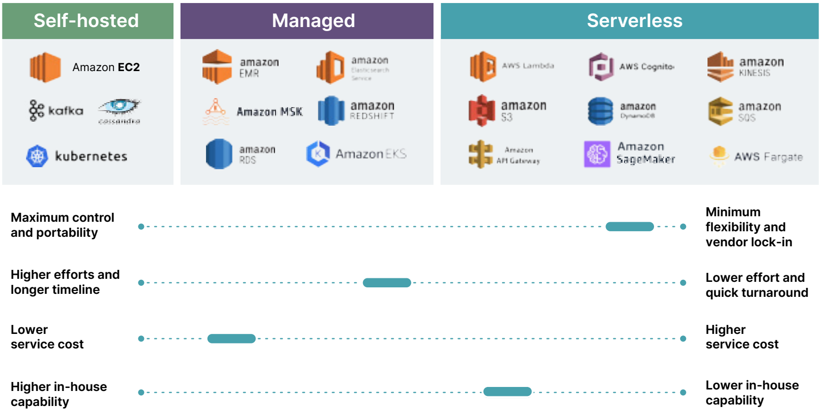
Managed services enable the cloud vendors to provide the expertise and automation to provision and manage the operational aspects of the service or application. Today, these include cloud services such as managed K8s cluster, like Amazon's Elastic Kubernetes Service (EKS), managed relational databases (RDS), managed Redis cache, managed OpenSearch and managed Kafka such as Amazon Managed Streaming for Apache Kafka (MSK. Only a few popular open source frameworks are available as managed services. Managed services provide ease of operations and cloud portability as well as being cost effective.

Pro tip: Many AWS managed services using Graviton (ARM-based) processors provide better price-performance ratios and reduced environmental impact compared to traditional x86 options. E.g. AWS MSK

Self-hosted vanilla services. A self-managed service is one that runs on a pure vanilla virtual machine, for instance running MariaDB on EC2, and is responsible for patches, fixes, backups, high availability, software upgrades, dependencies, network infrastructure, security and so on. With this choice, you get maximum flexibility with tech stack but at higher effort and with longer timelines. Cost-wise, self hosted is cheapest, however it requires significant operations overheads.

Pro tip: Self-hosting is best suited for organizations with strong in-house cloud infrastructure management expertise.

Cloud service offerings trade-offs from self-hosted to managed to serverless services



Source: Thoughtworks

Which service to use?

Craft your cloud deployment architecture by carefully selecting services that perfectly match your workload requirements while remaining cost-effective. You can consider a hybrid cloud approach for optimal cost and flexibility.

- **For rapid prototyping, short-lived or low usage workloads, variable or unpredictable load, leverage serverless services to minimize effort and expedite development.** While serverless can be costlier, its benefits outweigh the expense in these scenarios. However, be mindful of potential vendor lock-in.
- **For high-utilization workloads, choose cost-effective options like self-managed or managed services.** These offer more control and avoid vendor lock-in, but require in-house expertise for provisioning and maintenance.

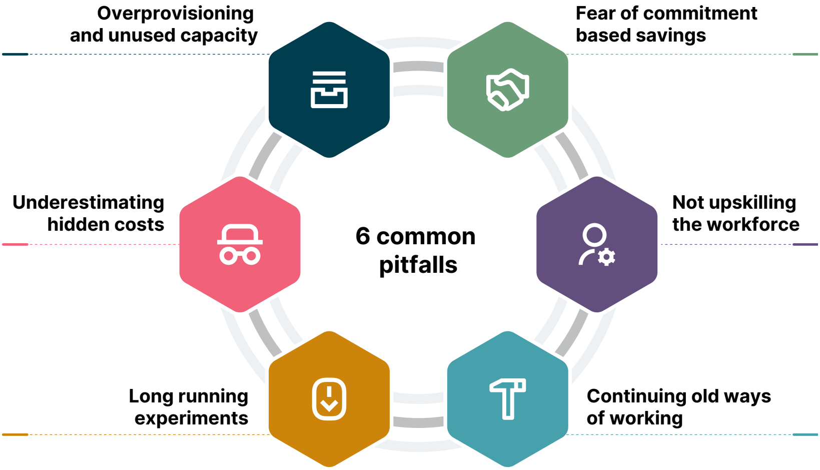
Pro tip: Strive for an 80/20 cloud service mix to optimize flexibility and minimize vendor lock-in. Build application architecture leveraging cloud-agnostic services for 80% of workloads. This promotes portability and reduces dependence on a single cloud provider. Allocate 20% of workloads to cloud-specific services for best value, ease and strategic use cases. Leverage the unique valuable capabilities offered by specific providers for specific use cases, while maintaining overall flexibility.

> **“A well-defined cloud service catalog with sensible defaults and best practices documentation is a fundamental component of a successful cloud operating model.”**

6

Six common pitfalls

The cloud offers immense flexibility and scalability, but it can also lead to unexpected costs if not managed properly. Here are some common pitfalls that can lead to unexpected cloud bills, and failure to realize the anticipated returns of cloud adoption. By being mindful of these pitfalls, and implementing best practices (covered in the next section: nine techniques of cost optimization), you can significantly reduce your cloud costs and maximize the value of your cloud investment.


Overprovisioning and unused capacity
Fear of commitment based savings
Underestimating hidden costs
6 common pitfalls
Not upskilling the workforce
Long running experiments
Continuing old ways of working

Overprovisioning and unused capacity

In the context of cloud costs, overprovisioning and unused capacity both lead to inefficient spending. For those coming from an on-premises world, it can be tempting to allocate more resources (compute, storage, etc.) than you need or use. It's like renting a mansion when you only need an apartment.

Common causes include:

- Over-provisioning for peak loads and static resource allocation.
- Lack of rightsizing and forecasting and ignoring utilization metrics.

Underestimating hidden costs

The underlying complexity of cloud service charges often leads to budget overruns. Unlike on-premises infrastructure, where network usage costs are generally not a concern, cloud environments introduce additional expenses that are frequently underestimated. Specifically, data transfer costs between Availability Zones (AZs), Virtual Private Clouds (VPCs) and regions can be substantial. And many enterprises have never encountered such costs in their on-premises environments, so don't factor them into their cloud budgets. This can result in unexpected charges and budgetary surprises.

Common causes include:

- Managed services extended support cost.
- Ignoring data transfer and networking cost.
- Not estimating observability cost, such as Cloudwatch.

Long-running experiments

With virtually unlimited on-demand capacity, teams can fall into the trap of running experiments indefinitely, or forgetting to terminate resources after use — much like paying for that fancy gym membership that you never use.

Common causes include:

- Never ending experiments, or forgetting to tear-down after experiments.
- Insufficient rigor about which experiments get funded.

Fear of commitment-based savings

Businesses often worry about committing to a specific level of spend or instance type for a one or three-year term, fearing they might underutilize resources if their needs change. Conversely, they might outgrow the reserved capacity, leading to additional spending.

Common causes include:

- Lack of forecasting due to uncertainty about future needs and poor capacity planning.
- Fear of vendor lock-in.

Not upskilling the workforce

A cloud-based environment requires different skills for managing and optimizing costs. Moving from manual on-premise process to fully automated on-demand cloud infrastructure requires coding skills and a different mindset.

Common causes include:

- Lack of capability building programs for upskilling the workforce.
- Not understanding cloud complexity and its pricing structure.
- Inefficient coding practices for Infrastructure as Code (IaC).

Continuing old ways of working

Migrating traditional IT practices to the cloud without adaptation often leads to inefficiencies, negating the cloud's primary advantage — its on-demand, elastic nature. Organizations that "lift and shift" old methods may end up over-provisioning resources for peak loads instead of leveraging auto-scaling. Additionally, manual provisioning and patching processes often result in "snowflake" servers (unique, difficult-to-reproduce configurations) rather than using standardized, immutable infrastructure. Manual operations result in configuration drift and increased complexity and costs.

Common causes include:

- Manual provisioning and patching processes leading to snowflake servers.
- Lift and shift migration strategy.
- Ticketed centralized manual process and lack of automation (IaC).

9

Nine cost optimization strategies

Drawing from our experience working with numerous large enterprises, we've developed a set of cloud cost optimization strategies designed to maximize cloud efficiency and minimize expenses.

Cost observability
Commitments
Rightsizing
Managed services
Private pricing and MAP
ARM Processors
Storage
Autoscaling
Networking and data transfer

Commitments with Savings Plans (SP) and Reserved Instances (RI)

Compute i.e. EC2 instances often dominate cloud bills. Cloud providers offer commitment-based discounts for these resources, potentially cutting compute costs by 30% up to 50%.

However, these savings apply only to compute resources, not other cloud services such as storage or managed services.

Follow three-step process to purchase SPs and RIs:

- **Analyze usage patterns.** Review your cloud resource usage over time to identify consistent, predictable workload patterns.
- **Choose the right commitment type.**
  - Savings Plans. More flexible, as they apply to multiple instance types/families across the EC2 instance families including EKS nodes, Lambdas as well as EMR Clusters.
  - Reserved Instances. Specific to an instance type/family, and Availability Zone. Best suited for managed services for databases such as RDS.
- **Regularly review and adjust:** Use cloud provider tools or third-party solutions to analyze commitment usage and readjust commitments by repeating these steps above. Remember to track plan expiration dates.

Start with a conservative commitment e.g., 50% of your baseline usage for a one-year term and no upfront payment option, and gradually increase commitments. Also purchase SP and RI at organizational billing accounts for flexibility during AWS account restructuring.

EC2 Instance	Memory GiB	vCPUs	On demand	Reserved cost (3 years, No upfront)	Savings
m5.large	8	2	$0.10	$0.04	56.44%
m5.xlarge	16	4	$0.20	$0.09	56.93%
m5.2xlarge	32	8	$0.40	$0.18	56.68%
m5.4xlarge	64	16	$0.81	$0.35	56.81%
m5.8xlarge	128	32	$1.62	$0.70	56.81%

Source: instances.vantage.sh (AWS Pricing for Asia Pacific Mumbai Region as of June 2024 with savings)

Pro tip: In case you end up committing to more than what is needed, leverage RI marketplace to resell your RI commitment. Read more here.

Antipattern: Avoid making large, upfront single purchases of Savings Plans or Reserved Instances as this can lead to overcommitment. Consider gradual multiple smaller purchases.

Additional resources: AWS Savings Plans, GCP CUD, Azure Savings Plan, Savings Plan vs Reserved Instances.

Rightsizing with effective resource utilization

Rightsizing resources is a crucial technique in cloud cost optimization because it ensures you're paying for exactly the computing power you need. Many cloud bills are inflated by overprovisioned resources due to lack of capacity planning and continuing old ways of provisioning resources. Rightsizing eliminates this waste by identifying and scaling down instances with low utilization. Based on experience, we have seen that rightsizing can often lead to significant cost reduction in the range of 30-70% of compute spend.

Near real-time monitoring of your cloud resources' utilization of CPU, memory, storage and network usage is essential for effective rightsizing. This monitoring provides valuable insights into usage patterns over time, including peak and off-peak periods. By analyzing this data, you can identify resources with consistently low utilization rates, allowing you to optimize their size and save on cloud costs.

Pro tip: Key to rightsizing is selecting the right instance family along with instance size. For example, in AWS choose R series instances for memory heavy workloads like databases and caches.

Pro tip: Use burstable T series and leverage Spot instances for experimental, non-critical and lower environments workloads.

Antipattern: Failing to regularly assess and adjust resource allocation based on actual usage can lead to inefficiencies. Rightsizing is an ongoing process that requires continuous evaluation and optimization, similar to fine-tuning performance.

Additional resources: Adidas Case Study, AWS Compute Optimizer, AWS Instance Types.

Autoscaling with horizontal dynamic scaling

The cloud's beauty lies in its scalability. With rightsizing, you can easily scale resources up during peak periods and down during low demand periods such as night time. This eliminates the need to pay for unused capacity.

Problem

Static provisioning for dynamic demand: Traditionally, cloud resources were provisioned with a fixed amount of compute power, such as CPU and memory. This static approach can lead to inefficiencies. Overprovisioning occurs when you're paying for unused resources during low-demand periods, a common issue given that most workloads experience fluctuations. On the other hand, underprovisioning can cause performance problems during peak demand, potentially resulting in lost revenue or a negative customer experience.

Solution

Dynamic scaling with horizontal autoscaling: To address these challenges, dynamic scaling through horizontal autoscaling offers a more flexible approach. This feature automatically adjusts the number of instances (virtual machines) by adding or removing them based on predefined metrics, ensuring that your application scales in real-time to meet demand.

Two autoscaling strategies: Metrics-based and schedule-based

- **Metrics-based scaling.** Based on defined thresholds for metrics such as CPU utilization or API Request traffic, when a metric breaches the upper or lower threshold, the autoscaler automatically adds or removes resources (e.g., virtual machines, containers) to meet the demand. **Metrics-based scaling is good for unpredictable workloads.** AWS EMR provides metrics-based managed scaling as part of managed services.
- **Schedule-based scaling.** Based on a defined schedule (e.g., weekdays vs. weekends, holiday season or sale for online retail, hourly variations) and the desired number of resources for each time period. The autoscaler automatically provisions or removes resources based on the schedule. **Scheduled-based scaling is good for predictable needs.**

Autoscaling automates resource provisioning, freeing up your team to focus on other tasks. Dynamic scaling with horizontal autoscaling is a game-changer for cloud cost optimization. Rightsizing pod resource requests in Kubernetes is also similar to rightsizing EC2 instances resources.

Pro tip: To achieve efficient dynamic scaling of nodes in EKS cluster, use tools like Karpenter that help not only with rightsizing the number of nodes, but also by choosing the best instance size for workloads deployed.

Antipattern: Relying solely on CPU utilization can lead to suboptimal scaling decisions. Consider using additional metrics such as network traffic, memory usage or custom metrics to get a more accurate picture of your application's resource needs.

ARM processors for general purpose workloads

ARM processors deliver the best price performance for general purpose applications. ARM processors in each CSP are Graviton for AWS, Axion for GCP and Ampere Altra for Azure.

Pricing of ARM based processors is significantly lower than Intel processors, however, not all software is currently optimized for ARM architectures. You might need to ensure your workloads are compatible with ARM processors before migrating.

EC2 Instance	Memory GiB	vCPUs	On demand	Graviton savings
m7g.xlarge	16	4	$0.16	19.05%
m7i.xlarge	16	4	$0.20	
m7g.2xlarge	32	8	$0.33	19.05%
m7i.2xlarge	32	8	$0.40	

Source: instances.vantage.sh (AWS Pricing for Asia Pacific Mumbai Region as of June)

Pro tip: Leverage Graviton processors for AWS Managed services such as RDS, MSK, EMR Clusters for direct price-to-performance benefit of ~ 20%.

Storage optimization

Cloud storage offers a seemingly endless pool of space, but neglecting optimization can lead to significant and unexpected costs. Not actively monitoring and analyzing storage usage can make it difficult to identify areas for optimization.

Storage optimization techniques include: Migrating EBS volumes from gp2 to gp3 type; rightsizing your EBS-provisioned IOPS rate; cleaning up detached EBS volumes; configuring storage lifecycle management policies and implementing intelligent tiering; choosing the right S3 storage class, such as Glacier storage, for archival purpose; cleaning up excessive backup snapshots; and compressing and deduplicating data.

Pro tip: Having a clearly defined data policy and governance strategy can significantly reduce costs associated with cloud storage.

Pro tip: Consider gp3 volumes, as it offers baseline IOPS with the ability to burst for additional performance when needed. The difficulty in predicting IOPS needs can lead to the common pitfall of allocating more IOPS than your workload requires.

Antipattern: Large datasets are particularly vulnerable to inefficiency, with duplicate copies existing for different teams or purposes.

Additional resources: Amazon EBS

Managed services optimization

While managed services reduce operational effort and the need for in-house expertise, it can also introduce challenges of vendor lock-in and higher cost if not utilized effectively. As a sensible default, we recommend that you choose managed services for open source components such as Kubernetes, Kafka, Postgresql or Mysql RDS, Redis etc. Understanding the cost of AWS managed services is crucial due to the complexity of pricing, which involves a pay-per-use model, variable factors and managed service overhead. To save costs, consider using RI (Reserved Instances) and Graviton processor.

Pro tip: Avoid the unnecessary cost of extended support for managed services by upgrading to the latest version. For example, the cost of EKS extended support is significantly higher than standard support.

Antipattern: Sending observability data from CloudWatch to a custom self-hosted or SasS observability stack can incur additional costs of processing metrics as well as egress charges.

Additional resources: AWS EKS Version and Support

Networking and data transfer optimization to avoid hidden cost

Networking components and data transfer costs are the most underestimated cost during cost calculation of AWS resources.

Knowing the purpose, use and associated charges of networking components is key to avoiding unnecessary networking costs. Choosing the right Gateway service from options including AWS Transit Gateway, NAT Gateway or Internet Gateway can be difficult as these components have different hourly rates and data transfer charges.

Data transfer costs in the cloud can be complex. Many factors influence pricing, making it difficult to predict exact costs upfront. Leverage cloud provider cost tools like AWS Cost Explorer to monitor and analyze data transfer usage and costs.

Pro tip: Multiple availability zones equals higher availability, but is it really required? Assess the trade-off between increased availability and additional costs when implementing a multi-AZ or multi-region architecture. Leverage service mesh zone aware routing and cell-based architecture to reduce data transfer cost.

Antipattern: Avoid using NAT Gateway for inter-VPC and on-premises network connectivity, as it is much costlier than other options such as Transit Gateway or VPC Peering.

Private Pricing (PP) and Migration Accelerator Programs (MAP)

Migration Acceleration Programs (MAP) provide a one-time migration benefit along with a suite of free tools, services and expertise to streamline the migration process. These programs are designed to facilitate the transition of applications from data centers to cloud infrastructure until they are stabilized and optimized.

Since MAP is complex and frequently updated, staying informed and following all guidelines when claiming benefits is crucial. Similar migration programs exist for GCP and Azure as well.

AWS also offers Private Pricing Agreements (PPA) or the Enterprise Discount Program (EDP) to high-volume, long-term AWS users who commit to a certain spend over a specified period (usually one to three years). Private pricing offers discounts on your overall AWS bill over and above your savings plans and RI commitment based savings. If you have high, predictable AWS usage and are comfortable with a spending commitment, private pricing can offer significant cost savings.

Pro tip: Leverage certified AWS Migration Competency Partners like Thoughtworks that can help you with your migration journey. These partners have a proven track record, as well as the expertise to deliver large-scale migration projects. Find out more here.

Antipattern: Inflating projected future spending to get better private pricing.

Cost observability dashboards

Effectively managing cloud costs requires a proactive approach. Democratizing real-time cost observability dashboards provide valuable insights into your cloud spending, empowering application teams to make informed decisions and optimize costs. This practice allows you to:

- Identify cost trends over time and spot any sudden spikes that require investigation.
- Gain insights into how efficiently resources are being utilized, helping you identify potential areas for right-sizing or scaling down.
- Generate alerts for anomalies in cost, usage and utilization patterns.

Pro tip: Select a cost and utilization dashboard tool that aligns with your specific needs and offers the functionalities required. For example, does the tool provide detailed analysis of container workloads inside EKS cluster? Does the tool provide utilization analysis for EMR clusters running MapReduce workloads? Does the tool provide recommendations for optimization?

Antipattern: Investing in a tool without understanding if the tool supports dashboards or recommendations for the type of workload running.

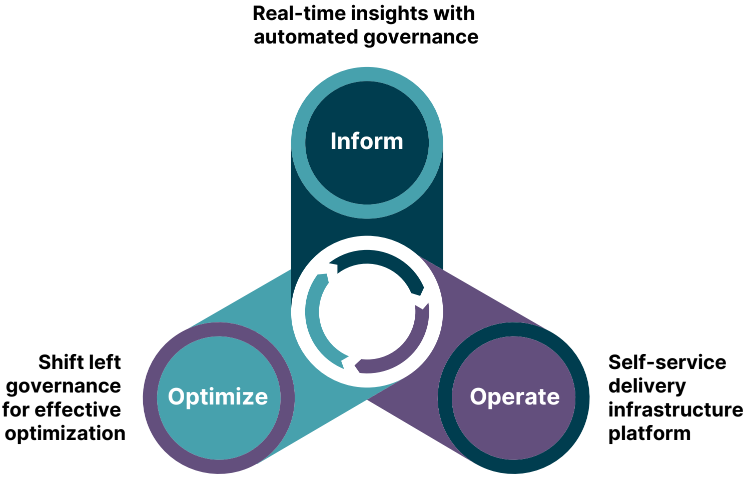
Additional resources: FinOps Tools

Disclaimer: This list of cost optimization techniques is not ranked, so be sure to leverage them based on your unique cloud service usage patterns.

3

Three pillars of effective cloud governance

Cloud governance is essential for organizations to ensure that their cloud infrastructure is used effectively, securely and efficiently. By implementing effective cloud governance, organizations can maximize the benefits of cloud computing while minimizing risks and costs. The following three pillars of cloud governance are aligned to the three phases of the Finops framework.


Real-time insights with automated governance
Inform
Shift left governance for effective optimization
Optimize
Operate
Self-service delivery infrastructure platform

Three pillars of cloud governance aligned to three phases of FinOps framework

Real-time insights with automated governance (Finops: Inform phase)

Empower everyone to understand infrastructure and application health by providing dashboards with utilization, performance and spending data.

A few recommended dashboards related to cost and utilization are:

- **Spend breakdown** transparent cost visibility at the application, environment and cloud service levels, including chargebacks for shared services, SPs and RIs, support fees, and accounting for MAP Credits and Enterprise discounts.
- **Utilization (usage) report** of CPU, memory and disks across all cloud services (EC2, EBS, EKS, RDS, EMR) and resources breakdown by application, environment and microservices running within Kubernetes.
- **Operational health dashboard**, showing performance-related data and the status of security, sustainability and compliance as part of the overall health of an application.

Implement ‘policy as code’ to automate the enforcement of compliance, security standards and recommended practices. Implement fitness functions with alerts to ensure ongoing adherence and proactively identify and address potential risks. For example, write a fitness function to Amazon EKS clusters version and alerts to be generated for running unsupported versions. To ensure swift action, clear guidelines should be established for non-compliance.

Pro tip: Implement fitness functions using serverless functions like AWS Lambda and push data to relevant systems with actionable alerts.

Pro tip: Cloud-agnostic tools offer the flexibility to monitor both your on-premises infrastructure and cloud environments seamlessly. This enables comprehensive observability in a hybrid setup without bias or limitations.

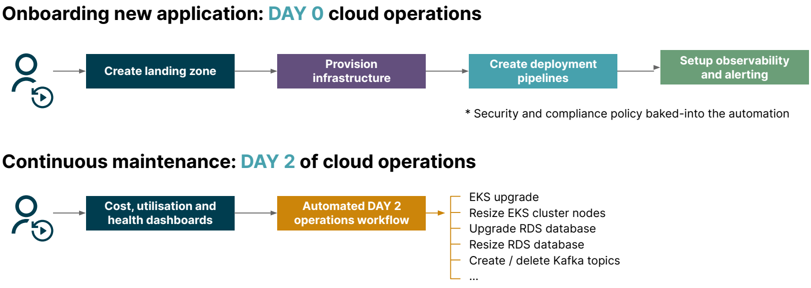
Additional resources: FinOps Tools

Invest in a self-service delivery infrastructure (DI) platform (Finops: Operate phase)

Self-service delivery infrastructure (DI) platform empowers developers and IT teams to provision, configure and manage their infrastructure environments (development, testing, production) on-demand without relying on manual intervention from an infrastructure team.

By investing in a self-serving capability with best practices baked-in, you can achieve faster deployments, improve consistency and reliability, and empower teams to be more efficient while operating at large enterprise scale. A self-service platform also helps with avoiding Shadow IT. Follow practices such as infrastructure as code and application templates while building a delivery infrastructure platform.

Self-service delivery infrastructure (DI) platform workflows



Onboarding new application: DAY 0 cloud operations

Create landing zone → Provision infrastructure → Create deployment pipelines → Setup observability and alerting

* Security and compliance policy baked-into the automation

Continuous maintenance: DAY 2 of cloud operations

Cost, utilisation and health dashboards → Automated DAY 2 operations workflow →

- EKS upgrade
- Resize EKS cluster nodes
- Upgrade RDS database
- Resize RDS database
- Create / delete Kafka topics
- ...

Pro tip: Leverage open source frameworks like Backstage.IO to build a Delivery Infrastructure Platform.

Shift left governance for effective optimization (Finops: Optimize phase)

Running pillars one and two manually and centrally at enterprise scale would be highly challenging. To streamline operations, learnings from how CSP operate suggest that these processes should be made self-service and democratized to application teams.

FinOps culture helps break down silos between finance, business and engineering teams. It also emphasizes aligning cloud spending with business goals such as measuring unit costs such as cost per customer, feature and team. It's not just about reducing costs, but also about optimizing resources to maximize the value delivered for the budget allocated.

Pro tip: By applying a rigorous tagging strategy to all cloud resources and accounts, we can achieve precise cost allocation, in-depth spend analysis, and actionable alerts to drive cost savings.

This approach ensures informed decision-making, efficient operations and ongoing optimization for a cost-effective and secure cloud environment.

Additional resources: FinOps Framework: Bringing accountability to cloud spend.

Metrics and KPIs for successful cloud cost management

Effective cloud cost management goes beyond just saving money. It's about optimizing your cloud resources to achieve the best value for your business. Here are some key metrics to measure success:

Metrics / KPI	Description
Cloud spend vs. budget (by application / project and environments)	Track your total cloud spend against your allocated budget. This helps identify areas of overspending and potential for cost reduction. Categorize applications / projects based on spend, e.g. S / M / L / XL and have automated daily, weekly and monthly spend tracking with overrun alerts. **Healthy state:** 100% of the cloud resources are tagged by application. < 2% budget overrun, with quarterly budget revisions.
Spend by services by application and environments	Analyze the cost breakdown of individual services by environments running in the cloud. This helps pinpoint areas where resources might be underutilized or highly expensive and allows you to look for alternatives. Monitor spend ratio from Prod vs NonProd environments. **Healthy state:** More than 80% of workloads are running with committed SP and RI. (Utilization of SPs and RIs should be 100%).

Metrics / KPI	Description
Resource utilization (CPU, memory and disk)	Monitor the average utilization of CPU, memory and storage resources across your cloud instances. Low utilization indicates potential overprovisioning and wasted spending. Cloud instances also include EKS Nodes, EMR Cluster Nodes and Containers like Pods running inside K8s cluster. **Healthy state:** Ideally, EC2 instances should run above 70% utilization in cloud infrastructure, leveraging autoscaling to cater for peak loads and festive seasons.
ARM processor adoption percentage	ARM processors such as Graviton help reduce cost as well as carbon footprint, leading to green sustainability initiatives. By default, use Graviton for MSK, RDS services. **Healthy state:** Having more than 25% of workloads running on Graviton with quarterly targets of a 2% increase.
Observability coverage	Ensure gathering of observability data from all relevant sources, including applications, infrastructure components and network traffic. Look at all three aspects of observability - metrics, logs and traces. Also monitor the cost of observability infrastructure as compared to cost of application hosting. **Healthy state:** Ensure more than 95% resources / applications are observed well. In best cases, the cost of observability should be in the 10-15% range, with the worst case not exceeding 25% of the application hosting cost.

3-6-9-3 Strategy of optimal cloud economics

> **“The cloud is transforming how businesses operate, requiring enterprises to adapt to a new model for consuming and managing infrastructure resources.”**

For optimal cloud economics, it’s essential to understand the three types of cloud services and their trade-offs, avoid six pitfalls that lead to an increase in cloud spend, and implement nine techniques for cost optimization by adhering to the three pillars of effective cloud governance.

The cloud landscape is constantly shifting, with new services and best practices emerging all the time. To stay ahead of the curve, it’s crucial to continuously learn about these advancements and their cost structures. By embracing migration as the default (even from one cloud service offering to another), organizations can leverage the numerous benefits of the cloud and gain a competitive edge.

Understand	**Avoid**
3	**6**
types of services	common pitfalls
Apply	**Uphold**
9	**3**
optimization strategies	pillars of governance

Author

Sunit Parekh
Head of digital platforms practice, Thoughtworks

With over 20 years of experience, I'm a seasoned technology strategist passionate about helping clients achieve their digital goals. I specialize in guiding large enterprises through complex distributed projects, from global solutions to digital transformations. My expertise lies in crafting impactful technology strategies and implementing cutting-edge cloud-native solutions across ambitious projects.

Modern engineering advocate and cloud-native champion
I'm a firm believer in leveraging the power of cloud ecosystems and embracing cloud-native approaches to build modern, scalable infrastructure. I'm equally passionate about collaborating with clients who share my commitment to adopting modern engineering practices for achieving technical excellence.

Open source contributor
Beyond my client work, I actively contribute to the open-source community. I've built a valuable tool, Data Anonymization, that helps developers safely anonymize production data for testing purposes.

Thoughtworks is a global technology consultancy that integrates strategy, design and engineering to drive digital innovation. We are over 10,000 Thoughtworkers strong across 48 offices in 19 countries. For 30+ years, we've delivered extraordinary impact together with our clients by helping them solve complex business problems with technology as the differentiator.

thoughtworks.com

/thoughtworks

Strategy. Design. Engineering.