FinOps Playbook 2025

Your Strategic Guide to Cloud Cost Optimization in Latin America

Reduce cloud spending by up to 40% and recover your investment in less than 90 days.

Quick Index

1. Executive Summary

The technological landscape in Latin America is at a tipping point. Cloud adoption is not an option—it's the engine of growth and innovation. However, this accelerated growth brings with it a critical challenge: spending control.

Exponential Regional Growth: The cloud market in Latin America will add $26.4 billion dollars between 2025 and 2029, with a compound annual growth rate (CAGR) of 17.2%.[1] This boom, driven by digital transformation and artificial intelligence, presents an unprecedented opportunity, but also a significant financial risk if not managed properly.

The Pain of Cost Overrun: 84% of organizations identify cloud spend management as their number one challenge, exceeding their budgets by an average of 17%.[2] This cost overrun is not just a figure in a report; it represents a direct loss of capital that could be reinvested in innovation.

The FinOps Maturity Gap: Visibility is the first step toward control, and most companies haven't taken it yet. Only 22% of organizations can correctly allocate more than 75% of their cloud costs to the corresponding teams or products.[3] This 78% of unallocated spending is a blind spot where inefficiency and waste thrive.

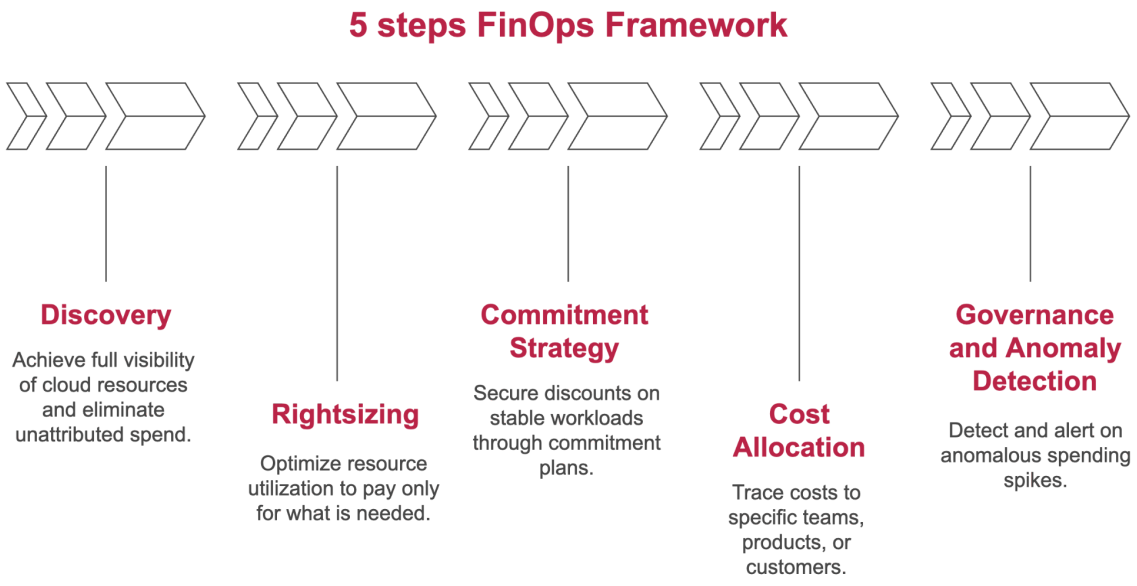
The New Frontier: Cloud+ and Gen-AI: The future is already here. The FOCUS 1.2 standard from the FinOps Foundation expands the scope of cost management beyond infrastructure (IaaS) to include SaaS, PaaS, and AI

token billing.[4, 5, 6] This demands multi-currency financial governance and a deeper understanding of digital services economics.

The h14z Promise: Our methodology, validated across leading organizations in Latin America, delivers typical savings of 20-35%, with peaks of up to 40%. We transform cloud spending from an unpredictable cost center into a strategic lever for profitable growth.

2. The 5-Step FinOps Framework: From Strategy to Execution

Our proven framework is a continuous cycle that aligns finance, technology, and business teams in a common language, turning cost data into intelligent decisions.



Below, we break down each step, combining strategic vision with tactical actions you can implement today.

Step 1: Discovery – Total Visibility, Zero Waste

Vision and KPI: Achieve 100% visibility of all cloud resources and their owners. The goal is to have not a single dollar of "dark" or unattributed spending.
Quick Win (≤ 7 days): Identify and eliminate orphaned resources (unattached disks, unassociated elastic IPs, old snapshots). This is the lowest-hanging fruit and can generate immediate savings of 5-10%.

Common Mistake: Ignoring SaaS and PaaS service costs like Snowflake, Databricks, or MongoDB. Cloud spending goes beyond virtual machines.[7]

Action Checklist:

☐ Perform an inventory of all resources in AWS, Azure, and GCP.
☐ Identify resources without owner or cost center tags.
☐ Map resource ownership to business units or product teams.
☐ Calculate current spending broken down by service and region.

Paso 2: Rightsizing Continuo y Automatizado

Vision y KPI: Achieve an average CPU utilization of 45% or higher across all non-scalable workloads. The goal is to pay only for what you actually need.
Multi-cloud Implementation: Use native tools from AWS, Azure, and GCP, plus solutions for private and hybrid infrastructure. Rightsizing is a universal principle.

Some examples in the main cloud providers:

Shell

```
AWS - Compute Optimizer
# Get EC2 instance recommendations
aws compute-optimizer get-ec2-instance-recommendations
# Get EBS volume recommendations
aws compute-optimizer get-ebs-volume-recommendations
# Get Lambda function recommendations
aws compute-optimizer get-lambda-function-recommendations
```

Shell

```
Azure - Advisor
# Get cost recommendations
az advisor recommendation list --category Cost
# Get VM-specific recommendations
az advisor recommendation list --category Cost --query
"[?contains(shortDescription.solution, 'virtual
machine')]"
# Export all recommendations to JSON
az advisor recommendation list --category Cost --output
json > azure_cost_recommendations.json
```

Shell

```
Google Cloud - Recommender
# Get Compute Engine instance recommendations
gcloud recommender recommendations list
--project=PROJECT_ID
--recommender=google.compute.instance.MachineTypeRecommen
der --location=ZONE

# Get disk recommendations
gcloud recommender recommendations list
--project=PROJECT_ID
--recommender=google.compute.disk.IdleResourceRecommender
--location=ZONE

# List all available recommender types
gcloud recommender recommenders list
```

Common mistake: Performing a "lift-and-shift" from on-premise servers to the cloud without resizing them. On-premise architectures are over-dimensioned by design; the cloud shouldn't be.[8]

Universal Principle: Rightsizing applies to any virtualized infrastructure, regardless of provider.

VMware vSphere: vRealize Operations, vSphere DRS recommendations
OpenStack: Tools like Watcher, Monasca for utilization analysis
Kubernetes On-Premise: Vertical Pod Autoscaler (VPA), tools like KubeCost
Híbrido/Multi-Cloud: CloudHealth, Flexera, Spot.io for unified visibility

Action Checklist:

☐ Implement CPU and memory monitoring on all critical instances
☐ Configure autoscaling policies for variable workloads
☐ Enable native recommendations: AWS Compute Optimizer, Azure Advisor, GCP Recommender
☐ For private infrastructure: configure vRealize Operations, Watcher (OpenStack), or KubeCost
☐ Deploy instance scheduler to shut down non-production environments outside business hours
☐ Establish target utilization metrics: 45%+ CPU for stable workloads

Step 3: Commitment Strategy (RI / SP / Spot)

Vision and KPI: Achieve an average discount of 40% or more on all stable and predictable workloads.
LATAM-Specific Consideration: Currency arbitrage is key. Before committing to 1 or 3 years, simulate the impact of foreign exchange (FX) fluctuation for currencies like MXN, CLP, BRL vs. USD. The new AWS region in Chile (expected by 2026) will open new latency and cost arbitrage opportunities for the Southern Cone.[9, 10]
Common Mistake: Over-committing to long-term savings plans just before an architecture migration or seasonal business peak. Analysis must precede purchase.

Action Checklist:

☐ Identify workloads with stable usage over the last 30-60 days.
☐ Use AWS and Azure calculators to simulate savings with different commitment types.
☐ Evaluate the use of Spot instances for fault-tolerant workloads, which can offer up to 90% savings.

Step 4: Cost Allocation – Kubernetes, SaaS and Beyond

Vision and KPI: Achieve more than 75% of all costs, including containers and SaaS, traceable to a specific team, product, or customer.[3]
Implementation (Code): Implement resource quotas at the namespace level in Kubernetes to establish limits and improve predictability.

JSON

```
# Example ResourceQuota for a team in Kubernetes
apiVersion: v1
kind: ResourceQuota
metadata:
  name: team-platform-quota
spec:
  hard:
    requests.cpu: "10"
    requests.memory: 20Gi
    limits.cpu: "20"
    limits.memory: 40Gi
```

Common Mistake: Mixing multiple products or services in a single namespace or account, making shared cost allocation impossible.

Action Checklist:

☐ Deploy open-source tools like OpenCost to visualize Kubernetes costs.
☐ Implement a Kubernetes labeling strategy by team, product, and environment.
☐ Integrate billing data from key SaaS platforms into your cost visibility system.

Step 5: Governance and Anomaly Detection

Vision and KPI: Detect and alert on any anomalous spending spike greater than 15% in less than 24 hours. The goal is to move from reaction to prevention.
Implementation (Policy as Code): Use tools like Open Policy Agent (OPA) to deny deployments that don't comply with tagging policies.
Common Mistake: Relying solely on alerts based on fixed budget percentages. An "80% of budget" alert is useless on day 28 of the month, but critical on day 3. Alerts must have a seasonal and contextual basis.

Action Checklist:

☐ Establish spending budgets and alerts in AWS, Azure, and GCP consoles.
☐ Create mandatory tagging policies for critical resources.
☐ Automate anomaly detection to identify deviations from normal spending.

3. Real Success Stories in the Region

Our methodology is not theoretical. It has generated measurable results for leading companies throughout Latin America.

Client (Anonymized)	Technology Stack	Project Duration	Key Result
Leading FinTech, Mexico	AWS + Kubernetes (EKS)	6 months	-60% cloud spending (savings of US $240K/year)

edTech, El Salvador	GCP + Cloud Run	2 months	-25% cloud spending

4. Key Implementation Templates

To accelerate your FinOps journey, here are some ready-to-use templates.

Minimal FinOps Team Structure

An effective FinOps team is cross-functional. You don't need a large team to start.

FinOps Lead (Business and Technology Liaison)
Cloud Architect (Technical Expert)
Financial Analyst (Budget and ROI Expert)
DevOps Engineer (Automation and Operations Expert)

Essential Tagging Strategy

A good tagging strategy is the foundation of visibility. Ensure every resource has, at minimum, these tags:

```json
{ "Team": "platform-engineering",
  "Environment": "production",
  "CostCenter": "ENG-001-PROD",
  "Owner": "nombre.apellido@suempresa.com",
  "Project": "nombre-del-proyecto"
}
```

Express Governance Checklist

☐ **OPA Policy:** "If Owner or CostCenter tag is missing, deny deployment".
☐ **Budget Alerts:** Configure automatic alerts at 50%, 80%, and 100% of monthly budget per team.
☐ **FinOps Review Meeting:** Schedule a mandatory monthly meeting with Finance and Technology leaders to review spending, savings, and upcoming projects.

5. Calculate Your Return on Investment (ROI)

Use this calculator to estimate your annual savings potential. Enter your current monthly cloud spending to see the impact of a well-executed FinOps strategy.

Optimization Tactic	% Typical Savings	Your Estimated Annual Savings (US$)
Idle Resource Cleanup	10-15%	
Instance Rightsizing	20-25%	
Use of Commitments (RI/SP)	30-45%	
Total Annual Savings	**20-35%**	

The payback period for an investment in FinOps consulting and tools is typically 2 to 4 months.

6. Next Steps: Your Implementation Roadmap

Week 0-1: Quick Wins	Month 1-2: Foundations	Month 3-6: Maturity
✔ Inventory and tagging	✔ Automated rightsizing	✔ Governance with Policy as Code (OPA)
✔ Orphan elimination	✔ Cost dashboards per team	✔ Optimized RI/SP/Spot strategy
✔ Basic budget alerts	✔ First RI/SP purchases	✔ Complete cost allocation (Full Cost Allocation)

Ready to get started? Book your free FinOps Assessment at https://calendly.com/h14z/finops-assessment

7. About h14z

We transform knowledge into competitive advantage with strategic AI.

We are a boutique consultancy specialized in Cloud FinOps and AI, with over US $2 million in demonstrated savings for companies in Latin America. We combine elite engineering, rigorous financial analysis, and automation to deliver measurable results in weeks, not quarters.

Contact: info@h14z.io

Schedule your Free Assessment: https://calendly.com/h14z/finops-assessment

Sources

[1] Cloud Computing Market in Latin America size to grow by USD 26.4 billion between 2025-2029 - Technavio Newsroom

[2] New Flexera Report Finds that 84% of Organizations Struggle to Manage Cloud Spend - Flexera

[3] The State Of Cloud Cost In 2024 - CloudZero

[4] Introducing FOCUS 1.2: SaaS/PaaS Support, Invoice Reconciliation, and Deeper Cloud Allocation - The FinOps Foundation

[5] FinOps Foundation's FOCUS 1.2 Expands to SaaS, PaaS - The New Stack

[6] FOCUS 1.2 Is What Finance Teams Have Been Waiting For - HyperFRAME Research

[7] State of FinOps Survey, 2025 - IDC

[8] Cloud Cost Challenges? Discover the Profitable Solutions to These 6 Common Challenges - Deimos.io

[9] In the works - AWS South America (Chile) Region - AWS News Blog

[10] Amazon to Invest More Than $4 Billion to Launch Infrastructure Region in Chile - Financial Post