Move2Cloud

• WHITE PAPER — 2027 EDITION

FINOPS & CLOUD OPTIMIZATION

Cut Your AWS Bill by 30 to 40%

The practical guide to combining pricing commitments, rightsizing and automation — and turning FinOps into a continuous discipline rather than a one-off audit.

11 pages · RI / Savings Plans / Spot comparison, self-assessment grid and roadmap included

29%
average cloud waste recorded in 2026

72%
max. discount with Reserved Instances

90%
max. discount with Spot instances

PREPARED BY
Move2Cloud — Cloud & Infrastructure Consulting, Paris

PUBLICATION
September 2026

CONTENTS

What you will find in this document

A six-stage journey, from understanding cost levers to implementing an operational FinOps roadmap.

01 **Executive summary — the essentials on one page** p.3

02 **Why FinOps Is Back on the Executive Agenda** p.4

03 **Structural Levers: Rightsizing, Graviton, Storage** p.5

04 **Pricing Commitments: RI, Savings Plans, Spot** p.6

05 **Automating Control: Scheduling, Autoscaling, Tools** p.7

06 **Illustrative Example: Combining the Levers** p.8

07 **Self-assessment grid** p.9

08 **A 5-step roadmap** p.10

09 **Conclusion & about Move2Cloud** p.11

> **Who is this document for?** CIOs, CISOs, FinOps leads and executives whose cloud bill is outpacing their forecasts, and who are looking for a repeatable method rather than a one-off audit.

Copyright Move2Cloud — move2cloud.fr

EXECUTIVE SUMMARY

The essentials on one page

After five years of continuous decline, cloud waste rose again in 2026 — and AI is the main driver.

According to the Flexera 2026 report, the average cloud waste rate has climbed back to **29%** of spend, the first increase after five years of steady decline. For a large enterprise spending more than 60 million euros a year on cloud — the case for 76% of the large organisations surveyed — that represents around **17 million euros of avoidable spend every year**. The main driver of this rebound: GPU instances, which cost 5 to 10 times more per hour than standard CPU instances, and for which legacy FinOps tools have not yet fully adapted their controls.

The encouraging sign: the FinOps discipline itself is progressing fast. According to the **State of FinOps 2026** from the FinOps Foundation, 98% of organisations now actively manage their AI costs, compared with 63% in 2025 and only 31% in 2024. But governance is struggling to keep up: only 28% of organisations have reached automated optimisation maturity.

The good news for teams that have not yet structured their approach: the classic levers remain powerful. Combined — pricing commitments (up to 72% discount with Reserved Instances, up to 90% with Spot), continuous rightsizing, storage clean-up and automated management — they **typically make it possible to document a 30 to 40% reduction in the bill over twelve months**, without degrading performance or availability.

29%
average cloud waste rate in 2026, rising for the 1st time in 5 years (Flexera)

€17M
avoidable spend per year for a typical large enterprise at the current waste rate

28%
only of organisations have mature automated FinOps governance

For an organisation using AWS — a growing startup, a mid-sized company or a large enterprise — the question is no longer "should we do FinOps?" but "which levers should we pull first, and how do we make this a continuous practice rather than an annual audit?". This document offers a pragmatic method to answer it.

Copyright Move2Cloud — move2cloud.fr

SECTION 02

Why FinOps Is Back on the Executive Agenda

For five years, cloud waste steadily declined as companies structured their FinOps practices. That cycle reversed in 2026, and three developments explain why the question now reaches the executive committee.

1. 1. AI has changed the scale of costs

A GPU-equipped instance costs 5 to 10 times more per hour than a standard CPU instance. Almost all companies now use generative AI in one form or another, and 58% of large organisations run generative AI workloads in production — up from 50% a year earlier.

2. 2. Visibility remains the main blind spot

Many FinOps teams report limited visibility into the real drivers of their AI costs, making it hard to measure return on investment or identify which teams generate the most spend. Without that visibility, it is impossible to prioritise optimisation levers with confidence.

3. 3. The governance gap is widening

Only 28% of organisations have reached mature, automated optimisation governance; most remain at an intermediate stage, with no controls specific to GPU workloads. The result: waste grows faster than teams can detect and fix it manually.

> *“ FinOps is no longer an audit you run once a year — it is a continuous practice, just like security or reliability. ”*

For any organisation whose AWS bill is growing faster than actual usage, this boils down to a simple question to ask today: do you know exactly where every euro goes, and who in your organisation is accountable for knowing?

SECTION 03

Structural Levers: Rightsizing, Graviton, Storage

Before negotiating a single pricing commitment, three structural levers are often enough to eliminate the most obvious waste.

LEVER	WHAT TO DO IN PRACTICE
Continuous rightsizing	Monitor instance utilisation continuously — not during a one-off audit. Sustained utilisation below 40 % over 14 days is a reliable signal of oversizing that needs correcting. Automate this assessment with AWS Compute Optimizer rather than handling it manually.
Graviton instances	For compatible Linux workloads, Graviton instances offer a better price/performance ratio than their x86 equivalents, often with no application code changes for high-level languages.
Storage optimisation	Downgrade high-performance EBS volumes to cheaper types (gp3, st1, sc1) when performance does not justify them, delete unattached volumes, and apply lifecycle policies to snapshots.

Key takeaway: These three levers require no contractual commitment and no major architectural change — they simply remove pure waste. They are almost always the most profitable starting point, even before negotiating any discount.

Why one-off rightsizing is no longer enough

A rightsizing audit run once a quarter captures a snapshot that goes stale within a few weeks as workloads evolve. The most mature organisations treat rightsizing as a continuous process, driven by automatic thresholds rather than manual reviews — which fundamentally changes the pace at which waste is detected and corrected.

SECTION 04

Pricing Commitments: RI, Savings Plans, Spot

Once the obvious waste has been eliminated, pricing commitments are the most powerful lever — provided each option is properly matched to the right type of workload.

OPTION	MAX. DISCOUNT	HOW IT WORKS	BEST SUITED TO
Reserved Instances	72%	A 1- or 3-year commitment on an instance type and a region; all upfront, partial upfront or no upfront payment. Convertible RIs allow the instance type to be changed.	Stable, predictable workloads over the long term
Savings Plans	up to 72%	A commitment in euros per hour, applied automatically across instance families, regions and eligible services — more flexible than classic RIs.	Stable workloads whose technical composition may evolve
Spot Instances	90%	Unused AWS capacity, sold at a variable price depending on demand; AWS can reclaim it at short notice.	Fault-tolerant workloads: batch, CI/CD, rendering, stateless Kubernetes clusters

Common pitfall: Deploying Spot instances on stateful workloads without a recovery mechanism (unprotected statefulsets, queues without persistence) turns a 90 % saving into a production incident. Spot only pays off when combined with a genuinely fault-tolerant architecture.

How to arbitrate between the three

A pragmatic approach: cover the stable, predictable base (databases, core business services) with Savings Plans to retain flexibility; keep classic Reserved Instances for workloads whose instance type will not change over the term of the commitment; and push everything genuinely interruptible towards Spot — batch processing, build environments, stateless workers.

Copyright Move2Cloud — move2cloud.fr

SECTION 05

Automating Control: Scheduling, Autoscaling, Tools

Structural levers and pricing commitments only hold over time if they are driven by automation rather than manual reviews.

Shut down what serves no purpose at night

Stopping non-production environments (development, staging, demo) outside business hours almost halves their cost: shutting an instance down 12 hours a day cuts its cost by close to 50 %. Tools such as AWS Instance Scheduler automate this with no manual intervention.

Take autoscaling beyond CPU

Many autoscaling policies watch only CPU or memory, which badly under- or over-sizes the capacity actually needed. Metrics closer to real usage — request count, latency, queue depth — produce more precise and therefore more economical sizing.

Detect drift in hours, not in thirty days

Compute Optimizer delivers automatic rightsizing recommendations based on observed real usage, while Cost Anomaly Detection raises an alert as soon as cost drift falls outside usual patterns. An anomaly detection tool that alerts a team within hours is worth more than a monthly report revealing the drift thirty days too late.

AWS TOOL	ROLE IN FINOPS CONTROL
CloudWatch	Collects the utilisation metrics that feed rightsizing and autoscaling decisions
Cost Explorer	Historical spend analysis by service, tag and account — the basis of all FinOps reporting

Copyright Move2Cloud — move2cloud.fr

SECTION 06

Illustrative Example: Combining the Levers

To make these levers concrete, here is an educational example — a fictitious company spending €100,000 per month on AWS, with no structured FinOps practice.

The levers applied and their estimated contribution

Savings Plans on the identified stable workloads contribute – 22 %; Spot instances on batch and CI/CD workloads – 6 %; continuous rightsizing of underused instances – 5 %; storage cleanup (orphaned volumes, snapshots) – 3 %; and automatic shutdown of non-prod environments – 4 %.

≈ 35 % cumulative on €100,000 per month

Taken together, these contributions represent an illustrative cumulative reduction of roughly 35 % on a bill of €100,000 per month — about €35,000 per month, or some €420,000 per year, with no degradation of performance or availability.

> **Key takeaway:** This example is illustrative and educational — each lever's contribution never adds up mechanically in a real environment, and results vary widely depending on the workload profile. It illustrates an order of magnitude that is achievable by combining several levers, not a guaranteed outcome.

What makes this scenario realistic: none of the five levers depends on an infrastructure investment or an application rewrite. They are operational and contractual adjustments, activable within a few weeks for the first three, and on a continuous basis for the last two.

Copyright Move2Cloud — move2cloud.fr

SECTION 07

Self-assessment grid

Before launching an optimisation programme, start by assessing your current maturity against these six questions.

- [ ] **Do you know precisely what share of your bill is on-demand versus committed (RI/Savings Plans)?** Without this baseline visibility, it is impossible to prioritise where to negotiate a commitment.
- [ ] **Is rightsizing a continuous process, or a one-off audit?** A quarterly review captures a snapshot that goes stale within a few weeks.
- [ ] **Are your non-production environments shut down automatically outside business hours?** If the answer is no, this is often the quickest win to achieve.
- [ ] **Are you using Spot instances for your genuinely fault-tolerant workloads?** Batch, CI/CD and stateless workers are the best candidates for a 90% discount.
- [ ] **Do you know the cost of your GPU/AI workloads separately from the rest of your bill?** Without this separation, it is impossible to measure the return on investment of your AI projects.
- [ ] **Does an automatic alert notify a team in the event of abnormal cost drift?** Discovering a drift on the monthly bill means a month of undetected waste.

> **How to read your answers:** Three "no" answers or more generally indicate a savings potential close to the 30 to 40% mentioned in the executive summary — it is the natural starting point for a structured FinOps audit.

SECTION 08

A 5-step roadmap

A progressive approach, designed to show results from the very first weeks.

1. **Measure**
   Put in place a consistent tagging system (ideally aligned with the FOCUS standard) to attribute every euro spent to a team, a project or a workload.

2. **Eliminate obvious waste**
   Delete unattached volumes, shut down orphaned resources, and correct the most blatant over-provisioning identified by Compute Optimizer.

3. **Commit**
   Cover stable workloads with Savings Plans, and migrate fault-tolerant workloads to Spot.

4. **Automate**
   Roll out automatic scheduling of non-production environments and evolve autoscaling towards business metrics rather than CPU alone.

5. **Govern**
   Put in place anomaly detection, showback or chargeback per team, and a short monthly review rather than a heavy annual audit.

Copyright Move2Cloud — move2cloud.fr

CONCLUSION

FinOps: a reflex rather than a project

Organisations that treat FinOps as a continuous practice — on a par with security — document sustainable savings of 30 to 40%, with no compromise on performance or availability.

With the rise of AI and GPU workloads, the stakes will only grow. Starting with visibility and the structural levers, before negotiating any commitment at all, remains the soundest foundation whatever your starting maturity.

Move2Cloud	is a consulting firm specialised in infrastructure, cloud and IT transformation. We support organisations in the audit, the optimisation and the securing of their cloud environment.
Contact	Move2Cloud — Paris · move2cloud.fr
Next step	An initial FinOps audit answers the six questions on page 9 within a few weeks, with a prioritised and costed roadmap.

Sources

ProsperOps, "Amazon EC2 Cost Optimization Guide for 2026" — Flexera, 2026 State of the Cloud report (summary via algeriatech.news, "Flexera 2026: Cloud Waste Returns — AI GPU Is Why") — FinOps Foundation, "State of FinOps 2026 Report" (summary via usu.com, "6 Takeaways from the State of FinOps Report 2026") — AWS documentation (Savings Plans, Reserved Instances, Compute Optimizer, Cost Anomaly Detection)

Document written by Move2Cloud as an example and illustration, for educational purposes. The savings percentages presented in section 06 are illustrative; actual results vary with each organisation's workload profile. This document is not a contractual commitment to results.