

NIVAAN LTD · KNOW WHAT YOU SPEND

Why Your Cloud Savings Never Happen

The gap between finding waste and actually recovering it, and how to close it in thirty days

Your dashboards are probably fine. Roughly half of everything they identify is never acted on. This is about the operating process between detection and someone doing the work, which is where the money is lost.

Version 1.0 · August 2026 · Nivaan Ltd · cg@nivaan.co.uk

1. The problem is not that you cannot see it

Most organisations arrive at cloud cost management assuming the problem is visibility. They buy a tool, or turn on the free one their provider already gives them, and wait for the savings.

The recommendations appear. Right-size this instance. Delete that orphaned volume. Buy commitment coverage for these stable workloads. The tool is working exactly as advertised.

And then, in most organisations, roughly half of those recommendations are never actioned. Not rejected on the merits. Not deferred with a reason. Simply never picked up by anyone.

That is the single most important number in cloud cost management, and it is not a tooling number. It is an operating model number. The difference between an organisation that saves five per cent and one that saves thirty per cent is almost never the quality of its detection. It is whether a recommendation becomes a piece of work that someone owns, with a date.

The test

Pick any recommendation currently sitting in your cost tool. Ask who owns it, when it is due, and what happens if it is still there next month. If you cannot answer all three, you have a visibility tool and no operating process. That is the normal state of affairs, and it is the thing this guide addresses.

2. Five reasons recommendations die

Understanding why the gap exists tells you where to intervene. In practice it is nearly always one of these five, and usually several at once.

It never becomes a ticket

A recommendation that lives in a dashboard requires someone to log in, read it, understand it, translate it into work, and enter it into the system where work is actually tracked. Every one of those steps is a place it stops. Engineering teams do not work from dashboards. They work from backlogs.

It has no owner

"The platform team should look at this" is not an owner. Cost recommendations that are addressed to a team rather than a person get treated the way a shared kitchen gets cleaned. If the tagging is incomplete, which it usually is, nobody even knows which team to address it to.

It competes with delivery and loses

An engineer who spends two days right-sizing instances and saves three thousand pounds a month has produced nothing visible in a sprint review. An engineer who ships a feature that adds five thousand pounds a month to the bill is congratulated. The asymmetry is structural, not cultural, and no amount of encouragement fixes it. What fixes it is putting cost work in the same backlog as everything else, sized and scheduled.

Nobody trusts the number

If tagging coverage is poor, the first response to any cost finding is to question the attribution. That argument is more comfortable than doing the work, and it can be sustained indefinitely. Trust in the data is a prerequisite for action, which is why tagging comes before everything else.

The feedback arrives too late to matter

A February overspend discussed in a March meeting is history. The architectural decision that caused it is three sprints old, the team has moved on, and unwinding it now costs more than it saves. Monthly reporting cadence is the wrong instrument for a cost that changes daily.

3. The fourteen day rule

One rule does more work than any tool you could buy.

Every cost finding becomes a ticket, with an owner and a fourteen day clock.

Not a dashboard entry. Not a line in a monthly report. A ticket in the system your engineers already work from, containing the affected resource, the current configuration, the recommended configuration, the estimated monthly saving, and the evidence behind it.

Fourteen days is not arbitrary. It is short enough that the context is still fresh and the resource still exists in the form described, and long enough to fit inside a normal sprint without disrupting planning.

What matters is less the specific number than the existence of a clock at all. A recommendation without a date is a suggestion, and suggestions lose to work that has one.

What a good ticket contains

Field	Why it matters
The specific resource	An identifier the engineer can search for. Not "the payments cluster" but the actual resource name.
Current and recommended state	So the change is a decision, not an investigation.
The saving, in pounds per month	Percentages do not motivate. A number does.
The evidence	Fourteen day utilisation profile attached, so the first response is not "are you sure?"
A named owner	Resolved from tags where possible. If tags cannot resolve it, that is the finding.
What breaks if done wrong	Honest risk assessment. Its absence is why engineers distrust cost tickets.

Six of the major cost platforms now create these tickets automatically into JIRA or ServiceNow. If yours cannot, that is a reasonable reason to change tools. If yours can and it is not switched on, that is a better afternoon than any procurement exercise.

4. Where the money usually is

The findings below recur across almost every estate. None of them require sophisticated tooling to detect. Most require a morning.

Committed and reserved capacity

The most valuable category, and the least visible, because it never spikes.

- Reserved instances and savings plans bought for workloads since resized, migrated or decommissioned
- Capacity sized against a load test that production traffic never reached
- Commitments purchased before right-sizing, which locks in the waste for one to three years

Anything reserved bills at full rate whether or not it is used. On a cost chart it draws a flat line, and every instinct says a flat line is healthy. Spikes get investigated. Flat lines get ignored.

Orphaned and idle resources

- Storage volumes left behind when their instances were terminated
- Snapshots accumulated over years with no retention policy
- Load balancers pointing at nothing
- Non-production environments running around the clock for teams who work office hours
- Secrets, DNS zones and certificates for services long since retired

Architecture that bills quietly

- NAT gateway charges on traffic that private endpoints would carry for nothing
- Services split across availability zones or regions, paying transfer charges on every internal call
- Debug level logging left on in production, billed per gigabyte ingested
- Support plans at enterprise tier on accounts with negligible spend

Attribution failures

- Untagged spend that cannot be assigned to any team or product
- Shared infrastructure with no allocation rule, so nobody feels responsible for any of it

Report untagged spend in pounds, not percentages

"Sixty-five per cent tag compliance" sounds like a passing grade. "Three point two million pounds a year of spend we cannot attribute to anyone" sounds like what it is. The metric you choose shapes the response you get.

5. Thirty days, no purchases

None of the following requires buying a product, running a procurement exercise or starting a programme. All of it can be done by people you already employ.

Week one: make it trustworthy

- Measure tagging coverage and express the gap in pounds per month, not as a percentage
- Block provisioning of untagged resources going forward. Preventing the problem costs thirty seconds; remediating it costs weeks
- Pull the reserved and committed capacity list with utilisation against each line, and identify who requested each one

Week two: make it visible weekly

- Replace the monthly report with a weekly message to each team containing three numbers: this week, last week, the difference
- Set anomaly alerts with meaningful thresholds. If every change triggers an alert, all alerts get ignored
- Shut down or resize anything on the capacity list running below meaningful utilisation

Weeks three and four: make it move

- Route every cost recommendation into your ticketing system with an owner and a fourteen day clock
- Schedule non-production environments off outside working hours, as the default rather than the exception
- Add cost estimation to your infrastructure pipeline so engineers see the monthly impact of a change before it merges, not months after
- Pick your largest workload and calculate what one unit of business work costs to run

If you do only one thing

Route recommendations into tickets with owners and dates. Everything else on this list is worth doing, but this is the one that converts detection into money.

6. Score yourself

Eight capabilities, scored Crawl, Walk or Run. One point, two points, three points. Total from eight to twenty-four.

Capability	Crawl	Run
Tagging coverage	Below seventy per cent, unenforced	Above ninety per cent, enforced at provisioning
Attribution	A material share cannot be assigned	Every pound maps to a team and a product
Action rate	Below thirty per cent of recommendations	Above seventy per cent within fourteen days
Ticketing	Findings live in a dashboard	Findings create tickets with owners and dates
Cadence	Monthly report	Weekly numbers per team, alerts on anomalies
Committed capacity	Bought, never reviewed	Every line has an owner and a utilisation floor
Non-production	Runs continuously	Scheduled off by default, exceptions opt in
Unit economics	Cost is a monthly total	Cost per unit of business work, tracked over time

Score	Stage	Where to start
8 to 13	Exposed	The data is not trusted, so nothing moves. Start with tagging and the capacity list. Nothing else works until the numbers are believed.
14 to 19	Aware	You can see it and it is not moving. Your problem is the detection-to-action gap. Start with ticketing and the fourteen day clock.
20 to 24	In control	Findings become work reliably. Move to unit economics and to preventing waste at provisioning rather than detecting it later.

A four minute interactive version covering twelve capabilities is at assess.nivaan.co.uk. Nothing is stored and no contact details are required.

7. When it is worth getting help

Much of the above is a genuine do-it-yourself exercise, and this guide is written so that you can. Three situations tend to be the exception.

- The people capable of doing it are the same people delivering your roadmap, so it slips quarter after quarter and has done for years.
- Attribution is broken badly enough that reconstructing it needs sustained work nobody has time for.
- The number has to go to a board or a regulator, and an internal estimate of your own waste is a harder document to defend than an independent one.

How Nivaan works

Stated plainly, because the objection behind most polite refusals is not price. It is that nobody has capacity for another initiative.

What we need	What we do not need
Read access to billing and usage exports	Access to any production system
One hour with your executive sponsor	Customer, member or personal data
One hour with finance	Software to buy, install or integrate
One hour with your platform or cloud lead	An agent or collector in your estate
One hour with an engineer	A procurement exercise
Two weeks of usage logs	A project team or steering group

Six hours

Your organisation's total involvement across a four week review. Four conversations and one data request. Everything else happens off to the side and comes back as a finished report: a figure in pounds with the evidence behind it, a breakdown by team and unit of work, savings ranked by value against effort, and a ninety day plan your own team can execute.

Independence

Nivaan sells no cloud, no tooling, no licences and no managed services, and takes no commission from any vendor. There is no product behind the advice. A recommendation to keep doing exactly what you are doing is an outcome we are paid the same for.

About

Nivaan Ltd is an independent cost advisory practice for UK financial services, drawing on twenty years of programme and portfolio delivery at HSBC, Barclays, Lloyds Banking Group, RBS, Refinitiv, Capgemini and Sopra Steria.

cg@nivaan.co.uk · nivaan.co.uk

Waste and action-rate figures cited from FinOps Foundation, Gartner and Flexera published research. Nivaan Ltd is registered in England and Wales. This guide is general information, not advice on your specific circumstances.