

BEYOND THE INVOICE

Cloud Migration Was the Easy Half — The Bill Is the Hard Half

The governance vacuum at the heart of enterprise cloud economics

Giovanni Leonardi

May 2022



Cloud Migration Was the Easy Half — The Bill Is the Hard Half

Portfolio · No. 90

Originally written May 2022 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“It is asking finance to be accountable for a line item that engineering controls, and asking engineering to optimise something they are not measured on.”

1. The Eighteen-Month Surprise

The quarterly business review is three slides in when the CFO pauses. The cloud migration programme — two years underway, celebrated internally for decommissioning four data centres and compressing provisioning lead times from weeks to minutes — has delivered its latest consumption report. The infrastructure cost line, which the original business case projected would flatten and decline by eighteen percent over three years, has risen by thirty-two percent in the past twelve months. The room goes quiet in the particular way it does when a number nobody expected collides with a narrative everybody believed.

The pattern is by now familiar enough to have its own rhythm. Organisations migrate significant workloads to a public cloud provider on business cases promising reduced infrastructure costs, elastic scaling, and freedom from the capital expenditure cycle. For the first nine to twelve months the narrative holds — the programme team can point to decommissioned hardware, avoided capital purchases, and the undeniable speed of provisioning. Then the consumption invoices arrive in earnest, and the conversation changes.

What follows is not a budget problem. It is a category crisis. The organisation discovers that it has swapped a cost model it understood — capital expenditure governed by procurement cycles, depreciation schedules, and approval committees — for a consumption model that nobody owns. Every engineer with API credentials has become a purchasing agent. Every auto-scaling rule is a standing order with no ceiling. And the monthly invoice, when it arrives, reads like a telephone bill in a language nobody in finance speaks.

This is not, at its root, a tooling gap. It is the consequence of removing the friction that was silently performing governance — and building nothing in its place.

2. The Friction That Was Doing the Work

In the traditional infrastructure model, cost control was embedded in process. Not because anyone designed it that way, but because the physics of procurement demanded it. Buying a server took weeks. It required a capital expenditure request, a business justification, architectural review, and sign-off at a level commensurate with the spend. By the time the hardware arrived, multiple people had been compelled to ask whether the expenditure was justified and the capacity correctly sized.

None of this was efficient. Much of it was actively frustrating. But it performed a governance function so deeply embedded in the procurement cycle that nobody recognised it as governance at all. It was simply how infrastructure worked.

Cloud adoption removed this friction by design — that was the entire value proposition. The commercial promise of infrastructure-as-a-service rests on the elimination of procurement lead times, capacity planning horizons, and capital approval gates. An engineer can provision a hundred virtual machines before lunch and decommission them by close of play — or, more commonly, provision them before lunch and forget them entirely.

The cloud did not break cost control. It revealed that cost control had always depended on procurement friction, and friction was gone.

The organisations that migrated earliest and fastest were, paradoxically, the most exposed. They had optimised for speed of adoption without recognising that speed of adoption was also speed of spend. The cloud providers' pricing models — pay-per-use, elastic, granular to the second — were designed to make consumption frictionless. And frictionless consumption, in an environment where every team provisions its own infrastructure, is precisely the condition in which costs become uncontrollable.

3. The FinOps Scramble — Remediation, Not Innovation

The emergence of FinOps as a practice — and the considerable energy behind it over the past two years — is best understood not as innovation but as remediation. What the FinOps movement is building is the governance layer that cloud adoption was designed to bypass.

This is not a criticism. The work is necessary, and much of it is thoughtful. But we should be clear about what it represents: the belated recognition that financial accountability in a consumption-based model requires entirely new mechanisms, not merely the adaptation of old ones.

The practical challenge is that those mechanisms must operate at three levels simultaneously, and most organisations have, in my experience, addressed at best one:

- **Visibility** — the ability to see what is being consumed, by whom, and for what purpose. This is where most organisations begin and where most tooling investment has concentrated: tagging strategies, cost allocation frameworks, dashboards that disaggregate the monthly invoice into something intelligible. Necessary, but insufficient. Visibility without accountability is surveillance — it tells you what

happened without changing what happens next.

- **Accountability** — the assignment of cost ownership to the teams and individuals who drive consumption. This is where the cultural shift begins in earnest and where most organisations stall. In a world where infrastructure was procured centrally, the IT department owned the cost. In a cloud-native model, every product team, every squad, every engineer who writes a Terraform module is making spending decisions — whether or not anyone has told them so.
- **Governance** — the decision-making framework that connects visibility and accountability to portfolio-level investment choices. This is where FinOps meets portfolio management, and where the gap is widest. An organisation can have excellent dashboards and meticulous tagging and still lack any mechanism for answering the question: is this the right amount to be spending on this capability, relative to the value it delivers?

4. Unit Economics: The Missing Conversation

The most consequential gap in most organisations' approach to cloud cost is not the absence of a tagging policy or a reserved-instance strategy. It is the absence of unit economics.

Unit economics — the cost of delivering a single unit of business value, whether a transaction processed, a customer served, a report generated, or an API call completed — is the bridge between infrastructure spend and business outcome. Without it, cloud cost is an input metric with no output reference. Finance sees a number going up; engineering sees infrastructure doing its job. Both are right. Neither is useful.

The difficulty is genuine. Unit economics requires a mapping between infrastructure consumption and business activity that most organisations have never needed to build. In the traditional model, the mapping was crude but workable: a server ran an application, the application served a function, the depreciation cost of the server was allocated to the function. In a cloud-native, microservices architecture, a single customer transaction may traverse dozens of services, each consuming compute, storage, and network resources from shared, auto-scaling pools. Attribution becomes a genuinely hard problem.

But it is the right problem. An organisation that knows its cost-per-transaction can make informed decisions about optimisation, about architectural trade-offs, about when to invest in efficiency and when to accept higher unit costs for speed to market. An organisation that knows only its aggregate monthly bill is navigating blind.

5. The Accountability Mechanism: Showback, Chargeback, and the Space Between

The debate over showback versus chargeback models is, at its core, a debate about how seriously an organisation takes cost accountability — and how much disruption it is willing to absorb to get there.

	Showback	Chargeback
Mechanism	Costs visible to teams; no budget impact	Costs allocated directly to team budgets
Cultural assumption	Transparency drives self-regulation	Financial consequence drives behaviour
Works well when	Strong existing cost culture; modest scale	Cost accountability is a strategic priority
Struggles when	Costs are large and delivery pressure is high	Shared-service attribution is immature
Risk	Visibility without consequence; nothing changes	Disputes over allocation; teams penalised for growth

In practice, showback works when the organisation already has a cost-conscious culture and when the numbers are small enough that nobody's budget is threatened. It struggles in precisely the conditions where it matters most — large, fast-growing consumption driven by teams under intense delivery pressure and no financial incentive to optimise.

Chargeback forces the conversation. When an engineering team's budget includes its cloud consumption, the trade-off between an expedient architecture and an efficient one becomes a financial decision, not merely a technical preference. The cost of over-provisioned instances, of forgotten development environments, of architectures that scale horizontally when they should scale vertically, lands where the decisions are made.

The resistance to chargeback is not all unreasonable. It introduces accounting complexity, creates disputes over shared-platform costs, and can penalise teams for traffic growth that is itself a sign of business success. But the alternative — treating cloud costs as a central IT overhead while distributing spending authority to every team in the organisation — is a structural absurdity. It is asking finance to be accountable for a line item that engineering controls, and asking engineering to optimise something they are not measured on. The predictable result is what we see across the industry: costs rising at a rate that bears no relationship to business growth, and periodic cost crises that produce short-term savings and long-term nothing.

6. The Cultural Shift No Dashboard Can Deliver

The deepest challenge — and the reason this is ultimately a governance problem rather than a tooling problem — is cultural. Cloud cost accountability requires engineers to think of themselves as economic actors, not merely technical ones.

This is a larger ask than it appears. Engineering culture has spent decades valorising technical excellence as the primary measure of quality. An engineer who shaves fifty milliseconds from response time is celebrated. An engineer who reduces the monthly cloud bill by ten thousand pounds is, at best, quietly acknowledged. The incentive structures — promotion criteria, sprint goals, peer recognition — overwhelmingly favour

delivery speed and technical sophistication over cost efficiency.

The organisations making genuine progress are the ones where this norm is shifting: where cost is visible in pull requests, where architecture decisions include an explicit cost dimension, where "how much will this cost to run?" is a standard question in design reviews rather than an afterthought raised by finance three months after deployment. In these organisations, cost is not a constraint imposed from outside; it is a design parameter as legitimate as latency or uptime.

7. The Bill Is Not a Surprise — It Is a Diagnosis

We should resist the framing of cloud cost as a problem to be managed down. The monthly invoice, read properly, is the most honest document in the technology portfolio. It tells you where your engineers are spending their time. It tells you which services are scaling and which are idle. It tells you whether your architecture is efficient or profligate, whether your capacity planning is working or ornamental.

The cloud providers, it should be noted, have no commercial incentive to facilitate this reading. Their interest lies in consumption growth, and their optimisation tooling — genuine and useful as it is — operates within the assumption that you will continue consuming their services. The question of whether a workload belongs in the cloud at all, or whether the architecture driving its consumption is the right architecture for the business need, is not one the provider's cost console is designed to answer.

The migration was the easy half. The discipline of living in the cloud — of treating consumption as investment, engineers as economic actors, and the monthly invoice as a portfolio governance instrument — is the work that most organisations have barely begun. And it will not be delivered by a dashboard.