inclusioncloud
DIGITAL ENGINEERING

EDITION 4

Is the “All to the Cloud” Era Over?

Time for the cloud reality check for 2026.

Executive Summary

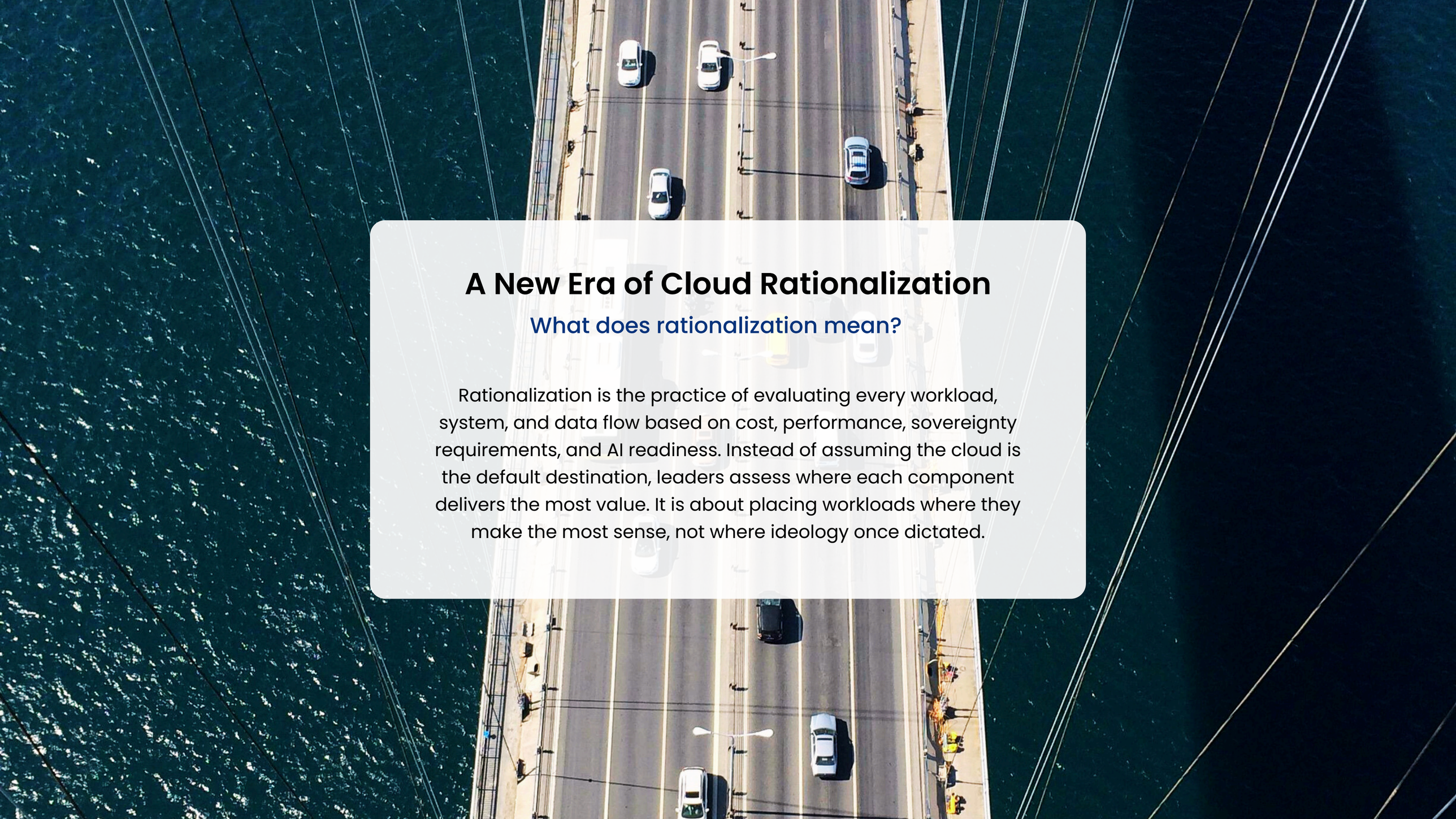
The Cloud Reality Check

For years, cloud migration felt like destiny. Every roadmap, every modernization initiative, and every board presentation pointed to the same place: upward.

Today the conversation is different. Boards and executives are not rejecting the cloud, but they are redefining their strategies. The push for 2026 is no longer a race to move everything to the cloud. It is about making smarter decisions regarding what belongs where. Public cloud. Private cloud. Hybrid. On premises. The mix matters now.

AI is a major driver behind this shift, but not the only one. Costs, sovereignty, data privacy, energy consumption, and performance are rewriting cloud strategy from the inside out.

This edition of AXIS explains why boardrooms are looking for a different path than the one they followed a decade ago and what this new direction means for cloud strategy.



A New Era of Cloud Rationalization

What does rationalization mean?

Rationalization is the practice of evaluating every workload, system, and data flow based on cost, performance, sovereignty requirements, and AI readiness. Instead of assuming the cloud is the default destination, leaders assess where each component delivers the most value. It is about placing workloads where they make the most sense, not where ideology once dictated.

For a decade the cloud was treated like the modern Rome. Everything led there. The assumption was simple: more cloud meant more modernization.

But that is no longer what boardrooms expect from their cloud strategy.

Executives are tapping the brakes on the “everything to the cloud” mindset and slowing down to ask sharper questions.

- → Which workloads actually perform better in the cloud?
- → Which ones cost less on premises?
- → Where data really needs to live?
- → And how AI can scale without setting the compute budget on fire?

This emerging paradigm does not mean CIOs are moving away from the cloud. What we are seeing is a more deliberate and disciplined approach to using it. We are entering a new rationalization era.

The fact is that the cloud is still central. Even more now, as companies begin to scale AI and the demand for scalable computing grows at an astonishing pace.

A few signals illustrate how organizations are adopting rationalization:

4 out of 5

cloud leaders plan to increase private cloud investments by 20% this year.

Source: Forrester

80%

of enterprises expected some workload repatriation in 2025.

Source: IDC

THE REALITY IS SIMPLE: **THE PUBLIC CLOUD IS AN OPTION AMONG OTHERS, NOT THE ONLY ROAD FORWARD FOR CIOS AND CFOS.**

1. The New Economics Behind Cloud Decisions

Cloud economics used to be relatively simple. Let’s look back for a moment: in the early days, the cloud was straightforward and predictable. It was a utility, just like electricity. You rented a generic server by the hour, and the math was linear: if your traffic went up, you added a server, and your bill went up a little. Easy. The big promise was simple, too: stop buying heavy hardware (CapEx) and just pay a monthly subscription (OpEx).

To be fair, complexity had been creeping in for years. As we moved from big, monolithic apps to thousands of microservices and containers, bills became harder to read. But even then, the logic remained the same: you paid for what you used, and resources were abundant.

AI was the tipping point. It didn't just add more line items to the bill; it changed the underlying physics of the market. We shifted from an era of abundance (cheap CPUs) to an era of scarcity (expensive GPUs) and extreme volatility.

This shift forces us to rethink three critical parts of our budget.

1.1 It’s not just about the bill anymore; it’s about visibility.

Hidden costs used to be rounding errors. Now, they are budget-killers. We are facing a phenomenon known as "Data Gravity."

Think of your data like a planet: the more you accumulate, the heavier it gets, and the stronger its gravitational pull becomes. It becomes incredibly difficult (and expensive) to move, effectively forcing you to bring your applications to the data. For AI, which feeds on massive datasets, this is a trap. If your data sits in one cloud but the GPUs you need are in another, moving that data triggers massive transfer fees.

1.2 Energy is no longer just a footnote. It's a boardroom topic.

For years, the public cloud offered a value proposition where energy was a fixed, internal cost that was easily absorbed. The sheer, intense power demand of AI workloads is now breaking that assumption.

Hyperscalers' massive data centers still benefit from economies of scale and access to renewable energy, which makes them far more efficient than local facilities. Yet, even they are starting to feel the strain. According to Bloomberg, electricity prices around major data-center hubs have surged by up to 267% in the last five years, and demand for grid power in the U.S. is projected to grow 22% in 2025 alone, adding roughly 11.3 GW to the system. (Pew Research Center)

The warning lights are starting to blink. When you buy cloud (public or private) the energy cost is already baked into your bill. But as Forrester analysts Lee Sustar and Charlie Dai point out:

> *Every public cloud customer is already funding this AI transformation. You are already paying either directly for managed AI services or indirectly through standard cloud bills. Those commodity services cost less to deliver now, thanks to billions saved by extending server lifespans, savings that hyperscalers are plowing straight into AI infrastructure.*

As AI's appetite for computation and cooling grows, those same bills could rise in ways few CFOs have yet modeled. At some point, hyperscalers may begin passing on their rising electricity costs to customers. On-premises setups, meanwhile, offer the illusion of control. In practice, measuring and managing real power usage can be a blind spot.

Electricity costs are often difficult to track, especially now that power-hungry AI workloads are new to many organizations. Training large models can drive consumption spikes that inflate utility bills and surprise financial planners.

The result is a new kind of tension in the C-suite. Every teraflop of compute has a wattage behind it, and every watt has a price. The cloud no longer floats above the financial discussion; it's plugged directly into it. Energy has quietly become the hidden currency of the AI era.

1.3. The "CAPEX vs. OPEX" promise is getting complicated.

Cloud economics were built on a clear distinction. Capital expenditures (CapEx) meant buying servers, building data centers, and carrying those assets on the balance sheet. Operational expenditures (OpEx) meant paying only for what you used in the cloud. That flexibility (the ability to scale up or down without long-term obligations) was the core promise of public cloud services.

In the early cloud era, this model worked exactly as intended. Compute was abundant, pricing was predictable, and enterprises could finally avoid the multi-million-dollar hardware purchases that defined traditional IT. “Pay as you go” really meant paying only for usage, not for capacity.

AI has redrawn this picture.

Modern AI workloads require high-performance GPUs, enormous datasets, and continuous pipelines of storage, networking, and data movement. CIOs quoted by CIO.com describe AI-heavy cloud bills as “astronomical,” noting projects that end up costing three to four times more once egress, data transfer, and the 10 to 12 supporting cloud services required for enterprise-grade AI are added.

Because GPU supply is constrained, cloud providers and infrastructure partners increasingly offer (and in high-demand scenarios, strongly encourage) multi-year capacity agreements to secure access to AI-ready compute. These contracts are still labeled OpEx, but many CIOs say they behave more like CapEx because they lock organizations into fixed financial commitments rather than true variable consumption. It is not universal across all workloads, but the pattern is visible enough that executives informally refer to it as “Shadow CapEx.”

Shadow CapEx: when OpEx spend behaves like CapEx

CapEx	Shadow CapEx	OpEx
rigidity/ asset	long-term fixed commitment, though accounted as Opex	flexibility/ consumption

The result is a new truth for CIOs:

With public cloud increasingly behaving like a fixed-cost model for AI rather than a variable-cost utility, FinOps becomes not just cost governance but financial risk management. The FinOps Foundation notes that mature disciplines can recover 10 to 20 percent of annual cloud spend and reduce waste by up to 30 percent, restoring some of the flexibility that enterprises originally associated with OpEx.

The cloud is still central, but its economics are no longer binary. It is not simply OpEx versus CapEx. It is a continuum shaped by AI demand, resource scarcity, and multitrillion-dollar infrastructure pressures.

A global AI-driven CapEx boom.

Business Insider reports that hyperscalers and AI leaders are investing at unprecedented levels:

- Amazon, Meta, Microsoft, and Google could spend $320 billion in CapEx this year alone.
- Meta expects to invest $600 billion through 2028.
- OpenAI and Oracle announced a $500 billion AI data-center project.
- Total global compute infrastructure may require $6.7 trillion by 2030, according to McKinsey.

This is one of the largest infrastructure expansions in modern tech history. That's why is called the “Capex boom”. And while these investments are made by hyperscalers, the financial ripple effects cascade to enterprise customers in the form of commitments, pricing constructs, and reduced elasticity, the very thing the cloud was supposed to avoid.

2. The Technical Side of the Story

2.1 The risk of vendor lock-in

Vendor lock-in has returned as a top strategic concern. As cloud platforms become more intertwined with AI services, the dependency on a small group of hyperscalers is increasing.

Gartner now classifies this dependence as a significant emerging risk. The concern is not only cost. It is the loss of strategic control. When business-critical workloads rely entirely on one provider, a single price change, outage, or compliance update can impact the entire organization.

In response, CIOs are adopting what analysts describe as "pragmatic" multicloud. The objective is portability rather than redundancy. This includes containerized applications, open APIs, shared data models, and Infrastructure-as-Code templates that allow workloads to move without major rework.

Even so, not every workload benefits from a multicloud approach. Some applications perform better in tightly integrated, single-vendor environments where unified tooling and high-performance networking improve speed and reliability. The goal for 2026 is not to eliminate lock-in entirely.

It is to choose it intentionally, using FinOps and architectural governance to understand where dependency creates value and where it limits future flexibility.

IN CLOUD AND AI STRATEGY, NOTHING IS FIXED. CIOS AREN'T CHOOSING A FOREVER-ARCHITECTURE, THEY'RE DESIGNING FOR ADAPTABILITY IN A LANDSCAPE WHERE THE BEST OPTION TODAY MAY NOT BE THE BEST OPTION TOMORROW.

2.2 Digital sovereignty and compliance take the spotlight

Cloud strategy is becoming dependent on geography. Governments are introducing new rules that define where data must reside, how it must be processed, and who may access it. For global enterprises, this elevates digital sovereignty from a regulatory requirement to a core architectural decision.

Digital sovereignty means preserving full control over organizational data and ensuring that it is governed under the laws of the region where the business operates. This requires companies to reassess workload placement, data residency, and provider selection based on local compliance capabilities.

Cloud providers are responding with features such as Confidential Computing, sovereign cloud regions, local processing zones, and AI-based threat detection. These capabilities allow organizations to keep sensitive workloads in the cloud while meeting strict governance standards.

The implication for CIOs is straightforward. Sovereignty is no longer a barrier, but it must be built into the architecture from the beginning.

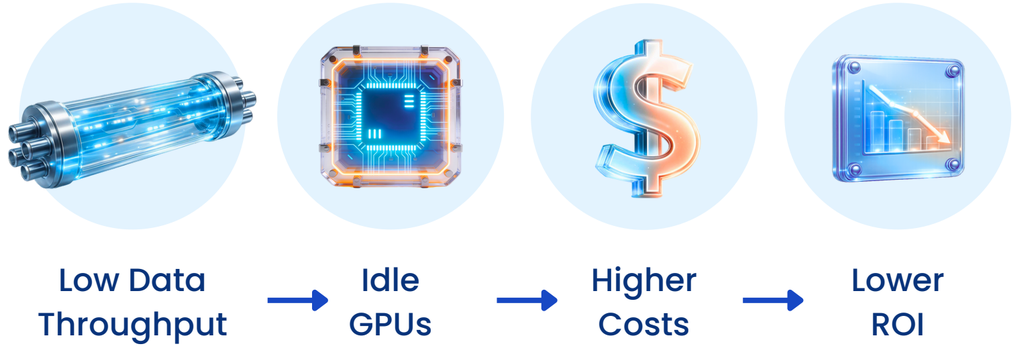
2.3 AI's technical bottleneck: throughput becomes the new battleground

As enterprises scale AI, a new technical limitation has emerged. Data throughput is becoming the defining constraint. When storage systems fail to deliver data quickly enough, GPUs remain idle. Idle GPUs increase costs and reduce the return on AI investments. This dynamic has become one of the most expensive inefficiencies within modern infrastructure.

In the early stages of cloud adoption, the priority was elasticity and rapid access to additional compute. In the AI era, performance depends less on the total compute available and more on the efficiency of data movement across hybrid environments. A slow data path can turn advanced GPU clusters into underutilized and costly assets.

To address this, organizations are modernizing the data layer as part of their cloud strategy. They are investing in NVMe-based storage, micro data centers, edge processing, and high-bandwidth interconnects that preserve proximity between data and compute. The benefits include lower latency, fewer egress costs, and significantly higher GPU utilization.

WHEN DATA CANT KEEP UP, GPUS GO IDLE



This trend also influences workload placement decisions.

WHEN PERFORMANCE, SOVEREIGNTY, AND COST PREDICTABILITY MATTER, MANY ENTERPRISES ARE CHOOSING TO RUN SPECIFIC AI AND DATA-INTENSIVE WORKLOADS ON-PREMISES OR IN PRIVATE CLOUDS WHERE THE ENTIRE DATA PATH CAN BE OPTIMIZED AND CONTROLLED.

3. Public, Private, Hybrid, or On Premises.

What works in 2026

After a decade in which the public cloud seemed like the only destination, new alternatives are opening up for executives. And AI has made this shift inevitable.

Not only have costs come under scrutiny, as we saw earlier, but the technical and strategic aspects of cloud use are being rethought as well. The goal is no longer to modernize for the sake of modernization, but to evaluate every factor carefully: ensuring data security, reducing latency, optimizing performance, controlling energy consumption, and maintaining regulatory compliance without sacrificing business agility.

That's one of the reasons behind the growth of what analysts call The Great Repatriation. It sounds dramatic, but it's not a return to the data centers of 20 years ago. It's a sign that companies are getting smarter about their architectures.

The motivations are clear: boards are demanding tighter control over costs, privacy, and compliance. And AI has amplified those concerns, not because the technology itself is flawed, but because it introduces new risk vectors.

90%

of organizations will adopt a hybrid cloud approach by 2027.

Source: Gartner

Many IT leaders worry about sensitive data being used to train public models or crossing regulatory boundaries. In highly regulated sectors, that's a line you can't cross.

Still, moving some workloads back on-premises doesn't mean giving up the flexibility that made the cloud attractive in the first place. The new generation of as-a-service offerings is bridging that gap, bringing cloud-like scalability and consumption models into private environments.

Hybrid is no longer a temporary stage. Gartner predicts that 90% of organizations will adopt a hybrid cloud approach by 2027, and the most urgent GenAI challenge to address over the next year will be data synchronization across hybrid environments.

4. The Human and Organizational Levers

The part of cloud we don’t talk about enough

Beyond the economic and technical factors, cloud migration is still a human challenge. These programs are complex. They require specialists who understand both the legacy environments being replaced and the modern architectures taking their place. But they also depend on something less visible and far harder to master: coordination.

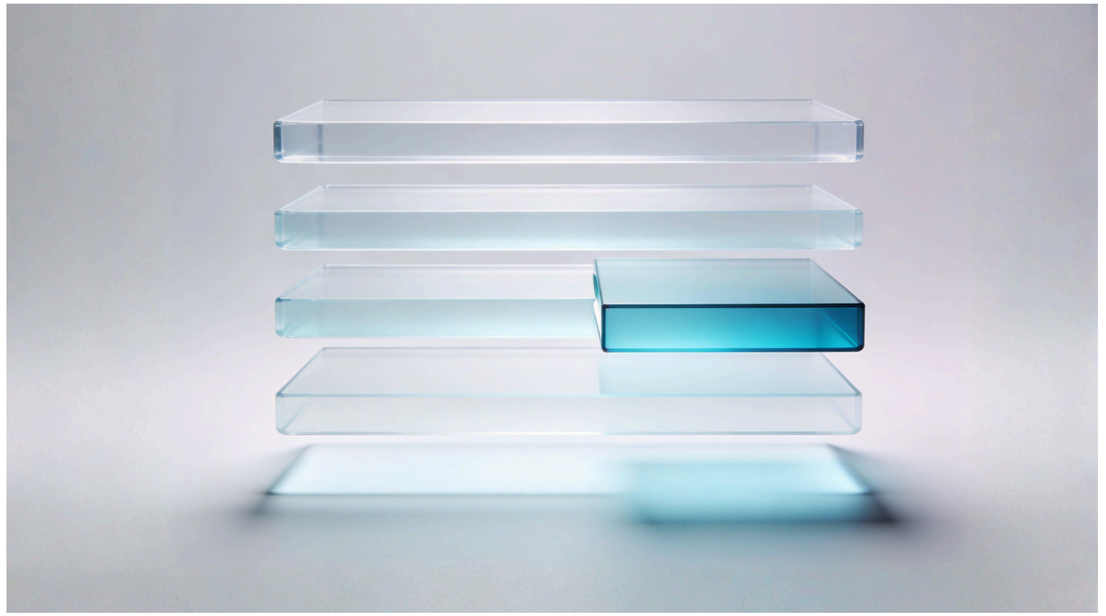
Most organizations don’t have all the expertise they need in-house. That’s why they rely on partners. And to get the most out of these collaborations, defining clear responsibilities and scopes of work becomes essential.

In the following sections, we’ll look at both sides of this equation: first, the growing cloud skill gap; and second, how companies are learning to coordinate more effectively with the partners who help close it.

50%

of respondents cited “talent & skills gaps” as the hardest barrier to digital sovereignty.

Source: InclusionCloud



4.1 The cloud skill gap widens

The demand for cloud expertise keeps growing, but the talent pool isn't keeping up. According to TechTarget, IDC estimates that more than 90 percent of organizations will face IT skill shortages by 2026, costing them around 5.5 trillion dollars worldwide. A significant portion of that gap lies in cloud skills.

And it's not slowing down. Gartner projects that global spending on public cloud services will hit 723 billion dollars in 2025, a 21 percent jump from last year. Yet many organizations can't find enough qualified professionals to manage that growth.

The consequences are everywhere. Projects stall halfway. Deployments stay unstable. Teams spend more time fixing problems than improving systems. And as TechTarget notes, many organizations are forced to delay innovation just to keep things running.

Cloud roles today go far beyond managing servers. Engineers now need to juggle multi-cloud environments, implement FinOps practices, and navigate complex compliance and digital sovereignty requirements. Add AI into the mix, and the pressure multiplies.

CLOUD ROLES IN HIGHEST DEMAND FOR 2026

Role	What They Focus On
Cloud Architect	Multicloud design, governance, cost-optimized architectures.
Cloud Engineer	Building and operating cloud infrastructure and automation (IaC).
DevOps / Cloud-Native Engineer	CI/CD, containers, Kubernetes, cloud-native delivery.
Cloud Security Engineer	IAM, compliance, threat defense, secure architectures.
FinOps Specialist	Cost optimization, cloud spend governance, usage insights.
Cloud AI/ML Engineer	Deploying AI models, data pipelines, and inference at scale.

4.2 Partner coordination becomes a strategic factor

But what happens when a partner joins the project to help close the skills gap?

Suddenly, more hands are in the mix, and that makes coordination even more complex. Without a clearly defined strategic plan, responsibilities can overlap, and critical tasks can fall through the cracks.

That’s where well-established frameworks come in. Tools like RACI, DACI, or RAPID can help map out who is Responsible, Accountable, Consulted, and Informed at each stage of the migration. These frameworks bring structure to collaboration and ensure every stakeholder knows their exact role and scope.

Final Insights

Finding Balance Beyond the Ideology

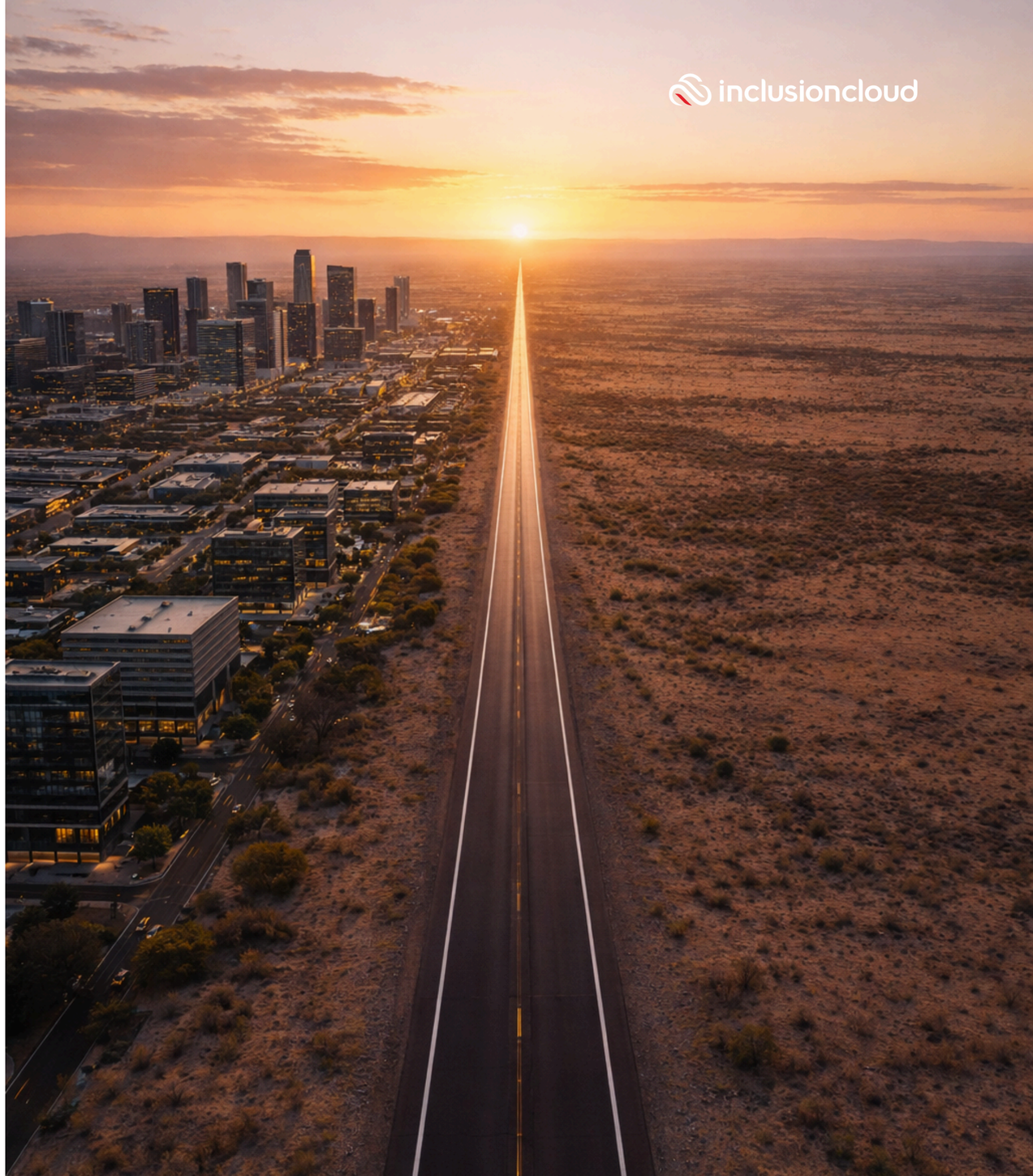
For years, the idea of moving everything to the cloud carried a sense of inevitability. It was the storyline everyone agreed on, even when the reality was messy. But by 2026, the gap between the narrative and the day-to-day operations has become impossible to ignore.

- Data that cannot leave certain regions.
- Models that cannot train on public clouds.
- Costs that refuse to flatten out.
- Architectures that remain as tangled as before.

Faced with this, CIOs and CFOs are abandoning ideology and shifting to what actually works. Sometimes that means repatriating workloads. Sometimes it means splitting vendors. Sometimes it means building a hybrid model that trades simplicity for resilience.

This is the real evolution: embracing complexity in order to get results. And as companies juggle multiple architectures, the demand for talent that can coordinate everything grows. The cloud solved many problems, but it also introduced new ones. In the end, 2026 is less about following a prescribed direction and more about choosing the path that produces the best results.

The winners will be the organizations that stay pragmatic, understand their own needs, and make architectural decisions based on value rather than shiny trends.



Disclaimer

The insights presented in this report draw from third-party research, industry publications, and Inclusion Cloud's own analysis of cloud, data, and AI initiatives. They reflect trends and perspectives observed across multiple sources but may not capture the full diversity of scenarios in every industry or region.

While we strive for accuracy and relevance, these findings should be interpreted within the context of the referenced sources and the scope of Inclusion Cloud's research. This report is for informational purposes only and should not be considered financial, legal, or architectural advice. Inclusion Cloud is not responsible for any decisions or outcomes resulting from the use or interpretation of this material.

About Us

Inclusion Cloud helps companies implement and optimize enterprise software, cloud solutions, data platforms, and AI capabilities. As certified partners of SAP, Oracle, ServiceNow, and Salesforce, we deliver the technical expertise and integrations that modern enterprises rely on.

Our AI-powered talent engine, inMOVE™ by Inclusion Cloud, combines advanced recruiting technology with rigorous technical validation to secure top certified resources. With more than 20 years supporting global organizations, we focus on building solutions that are scalable, integrated, and ready for the real world.

inclusioncloud

Discover how we're helping enterprises build what's next at inclusioncloud.com