Cloud Concentration Risk

When Efficiency Becomes Fragility

Author: Che-Hwon Bae
Website: www.baemax.co.uk
Version: V3 — Revised Working Paper
Last Updated: January 2026

Disclaimer

This paper presents an analytical framework for discussion. It does not constitute investment, legal, military, or policy advice, nor does it advocate specific actions. The views expressed are personal and intended to provoke informed debate.

Table of Content

Executive Summary

Over the last decade, cloud computing has delivered extraordinary gains in efficiency, scalability, and speed. Infrastructure that once required large fixed investment is now elastic, on-demand, and globally accessible.

But this shift has quietly reversed a principle that once underpinned system resilience: **diversification**.

As infrastructure has consolidated into a small number of hyperscale providers, the global economy has inherited a new form of concentration risk. What was once many organisations running their own imperfect but independent systems has become many organisations dependent on the same few control planes.

This note is not an argument against cloud computing. It is an argument that **efficiency has been mistaken for resilience**, and that concentration has introduced systemic risk that is poorly understood, weakly priced, and largely ignored until failure occurs.

The Original Trade-Off: Efficiency vs Resilience

Historically, organisations ran their own infrastructure.

It was:

- Expensive
- Inefficient
- Often under-utilised

But it had one critical advantage: **failure was local**.

A data-centre outage affected one firm, one region, or one sector. The blast radius was contained.

Cloud computing inverted this model.

Infrastructure became cheaper.
Scaling became trivial.
Reliability improved *on average*.

But the failure mode changed.

Failures became **correlated**.

Concentration Risk Disguised as Progress

Today, a large share of global digital activity depends on:

- A small number of hyperscale cloud providers
- A small number of network and traffic-routing layers
- A small number of shared authentication and orchestration systems

From the outside, everything appears robust:

- Multiple availability zones
- Redundancy within providers
- High headline uptime

But **diversification across providers is rare**, and genuine independence rarer still.

This creates a structural asymmetry:

> The system is robust to small failures and fragile to large ones.

Control Planes Are Single Points of Failure

The most dangerous concentration is not compute or storage — it is the **control plane**.

If identity, routing, orchestration, or access control layers fail:

- Redundancy inside the cloud does not help
- Failover logic itself may fail
- Human operators face opaque, cascading failure modes

In effect, much of the global economy now runs on **someone else's infrastructure — and someone else's control logic**.

This is efficient.
It is not resilient.

Scale Changes the Threat Model

Concentration does not just amplify accidents — it **amplifies incentives**.

As infrastructure consolidates, the payoff to compromise increases dramatically for:

- Malicious actors
- State-level attackers

- Insider threats

A successful control-plane failure or compromise:

- Affects thousands of organisations simultaneously
- Disrupts unrelated sectors at once
- Creates hostage-style dynamics rather than isolated outages

Security effort scales linearly.
Impact scales non-linearly.

More importantly, concentration reshapes attacker behaviour. As potential impact grows, so does the expected payoff from success. This attracts adversaries who are better resourced, more patient, and more sophisticated — including state-level actors willing to invest in long-term access rather than immediate disruption.

In highly concentrated environments, **espionage becomes at least as valuable as outage**. Persistent access to shared control planes offers visibility, leverage, and intelligence across thousands of organisations simultaneously. What appears efficient from an operational perspective increasingly resembles **strategic terrain** from an adversary's point of view.

Concentration turns infrastructure into a strategic asset — and strategic assets attract strategic adversaries.

The Illusion of "Someone Else's Problem"

Cloud adoption often comes with a psychological shift:

- Resilience is outsourced
- Responsibility feels transferred
- Tail risks are ignored

But while infrastructure is outsourced, **dependency is not**.

The organisation still bears:

- Operational risk
- Reputational damage
- Regulatory exposure
- Business interruption

The difference is that failure now arrives from outside the organisation — and often without warning.

When Shared Failure Creates Complacency

There is an often-overlooked psychological effect of large, correlated outages: **they create relief rather than urgency**.

When multiple firms fail simultaneously, responsibility diffuses. The fact that others are also down softens scrutiny and reduces the impulse to investigate deeply. Failure feels normalised rather than exceptional.

In these moments, organisations shift from asking *how could this have been prevented?* to *there is nothing we can do until it comes back*. Endurance replaces design.

This reaction is understandable. If an entire layer of infrastructure fails, individual firms have limited ability to intervene in real time. But the subtle danger is what happens afterwards. Shared failure suppresses learning. It reduces the incentive to question dependency, concentration, and architectural choices that made the failure possible.

Paradoxically, systems that fail rarely but catastrophically generate **less institutional learning** than systems that fail frequently but locally. Concentration turns failure into an external event rather than an internal design signal.

This mindset does not eliminate fragility — it normalises it.

What Would a More Resilient Model Imply?

If resilience is genuinely a priority, it cannot be delivered accidentally through efficiency. It requires deliberate design choices that break correlated failure.

In practice, this likely means **hybrid architectures** — not as a rejection of cloud computing, but as a recognition that not all system functions should share the same fate. Critical control, identity, and transaction-finality functions may justify higher fixed cost and operational complexity if they materially reduce systemic risk.

The trade-off is explicit: **higher baseline cost in exchange for smaller, more containable failures**.

That trade-off is uncomfortable — but pretending it does not exist simply transfers fragility elsewhere.

Structural Barriers to Diversification

Provider or failure-domain diversification is often discussed as a theoretical solution, but in practice it faces meaningful structural barriers. Proprietary service abstractions, data gravity, and egress pricing models all raise the cost of moving workloads or maintaining parallel environments. These frictions are not incidental; they reinforce concentration by making dependence economically rational even when it increases systemic risk.

As a result, organisations may recognise concentration risk while remaining locked into architectures that are difficult to unwind. This further supports the argument that the risk is weakly priced: market incentives favour efficiency and stickiness, while the cost of correlated failure is deferred and externalised.

As cloud infrastructure increasingly underpins critical economic activity, the distinction between commercial pricing and systemic lock-in becomes less clear. Exit friction — whether through proprietary abstractions or asymmetric egress costs — functions as a form of hidden leverage. Ensuring transfer and recovery paths are economically feasible may therefore be as important to resilience as uptime or security themselves, particularly where concentration risk has broader social consequences.

Why This Risk Is Under-Priced

Cloud concentration risk is rarely priced correctly because:

- Outages are infrequent
- Benefits are immediate
- Costs are delayed and socialised
- Incentives favour short-term efficiency

When highly concentrated infrastructure fails, the cost rarely remains confined to the firms directly affected. Disruption propagates outward to customers, employees, counterparties, and, in some cases, the public sector. End users absorb loss of access to essential services, while governments may face pressure to intervene where digital infrastructure underpins critical economic or civic functions.

In this sense, correlated failure can create implicit hostage-style dynamics. The more essential and concentrated the infrastructure becomes, the harder it is to allow failure to resolve cleanly through market mechanisms alone. The cost is not eliminated; it is redistributed — often to actors who neither chose the architecture nor directly benefited from its efficiencies.

This mirrors other systemic risks:

- Financial leverage
- Supply-chain concentration
- Energy dependency

They appear optimal — until they fail.

How Concentration Risk Could Be Priced

Today, cloud concentration risk is largely implicit. It is absorbed indirectly through business interruption losses, operational downtime, and reputational damage rather than priced explicitly at the infrastructure level.

In principle, however, this risk is not unusual. Financial systems already price correlated failure through mechanisms such as capital requirements, stress testing, and insurance premia. A similar logic could apply to infrastructure dependence. For example, insurance pricing that differentiates between single-provider and diversified architectures, or regulatory frameworks that require higher operational capital buffers for highly concentrated dependencies, would make concentration visible rather than assumed.

The absence of explicit pricing does not imply the absence of risk. It reflects a lag between architectural change and institutional adaptation — a familiar pattern in other systemic domains.

What Resilience Actually Requires

Resilience is not free, and it is not fashionable.

It requires:

- Deliberate redundancy
- Provider or failure-domain diversification
- Independent access and recovery paths
- Acceptance of higher baseline cost

These choices are often rejected as “inefficient”.

But **efficiency without resilience is borrowed time**.

Closing Thought

The shift to cloud computing did not eliminate infrastructure risk — it **repackaged it**.

We have traded many small, independent failures for fewer, larger, correlated ones. That trade may still be rational — but only if it is understood, acknowledged, and actively managed.

Ignoring concentration risk does not make systems modern.
It makes them brittle.

Once infrastructure becomes critical,
the ability to leave safely matters as much as the ability to scale quickly.