THE HIDDEN MIDDLE : WHY MID-MARKET CLOUDS WILL DEFINE 2026

The cloud's quiet revolution is happening between extremes.

TABLE OF CONTENTS

Introduction

Challenge

Transformation

Solutions

Citations

WELCOME TO THE
ELEPHANT IN THE ROOM!

You have in your hands a guide designed to call out the elephant in the room: a topic that's too important to be ignored but isn't getting the attention it deserves.

THE ELEPHANT

The story of cloud has always been told in extremes; hyperscale dominance versus sovereign defiance. But real transformation now lives in the middle.

The global cloud market is fragmenting. Hyperscale remain vital, but growth is changing as businesses rebalance workloads toward regional and private options. At the same time, sovereign clouds remain niche, serving tightly regulated verticals.

Between them lies **the hidden middle**: a segment poised for outsized growth. These are providers who combine modern infrastructure with proximity, compliance alignment, and predictable costs. They move fast enough for innovation yet stay close enough for trust.

As cloud services continue to grow at a projected 21.5% CAGR, regional and mid-market providers are positioned to gain share by differentiating on localization, regulatory alignment, and predictable costs—especially as data privacy regulations now apply in more than 79% of countries.[1,2]

INTRODUCTION

The Saturation Plateau

The major cloud migration is mostly finished. Most businesses now run hybrid setups across multiple providers. What started as expansion has turned into optimization.

CFOs and CIOs are asking tougher questions: Why are we still paying for unused capacity? Why are egress fees unpredictable?

The hyperscale promise of endless scalability conflicts with real-world cost pressures, data sovereignty rules, and performance inconsistencies.

In this environment, regional and mid-market providers have opportunities to stand out. They offer modernization without losing control and flexibility without excessive costs.

THE SATURATION PLATEAU

THE PERFECT STORM

The VMware Shake-Up and the Disruption by Design

The Broadcom acquisition of VMware has brought about significant changes for enterprises and businesses alike. As Broadcom continues to streamline the old VMware portfolio and resources operationally, more pressure is being passed on to VMware virtualization customers.

STEEP PRICE INCREASES AND HIDDEN COST SHOCK

Customers report that the switch to subscription-only models, bundled licenses, and "license-per-core" minimums has led to cost increases of up to **eight to 15 times**.[3]

LICENSING MODEL CHANGE AND COMPLIANCE RISK

The move from perpetual licenses and a broad SKU landscape to fewer offerings, higher minimum core counts, and a subscription-only model has forced customers into new contracts under tight timelines.[4]

PARTNER ECOSYSTEM DISRUPTION

Drastic changes in partner programs (from 4,500 to fewer than 100) and reseller chains have resulted in some customers losing their long-standing strategic partners or facing additional complexity when switching to an alternative.[5]

In 2025, VMware's largest 2,000 customers will shrink their deployment size by an average of **40%**.[6]

When Hyperscale Demands a
RESET

While hyperscale initially represented limitless capacity and cost efficiency, 2025 has shown us a different picture. With providers making significant investments in AI-focused data centers, enterprise clients are feeling the pinch of higher costs that offset these investments and experiencing steadily decreasing bargaining power, particularly for GPUs. Additionally, outages worldwide have made it clear how many organizations depend heavily on just a few hyperscale platforms.

The reality is, even as providers push forward with AI expansion, customers are finding their choices limited. This calls for a thoughtful rethinking of what a modern "cloud strategy" should really look like in the future.

Cost Management
Hyperscale cost reductions are often offset by higher consumption, and longer contracts may end up costing more for certain services.

Hyperscale Outages
Configuration changes, software updates, and cyberattacks continue to cause incidents, while their impact is becoming more severe globally—revealing the limits of scale and shared architecture.

Rising LLM Technology Cost
Premium pricing for AI and HPC specialized services drive up costs. After all, who is paying for the 50% increase in power needs by 2027 to train and productionize LLM?[7]
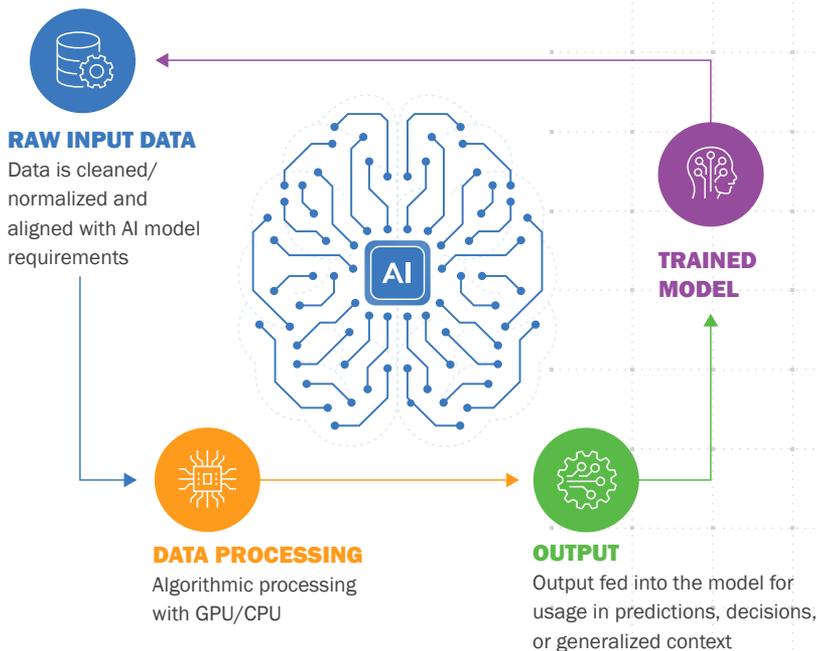
Reduced Customer Bargaining Power
Hyperscale purchases of land, power, and equipment (especially GPUs) provide significant bargaining power and reduce resources available for others.

Azure
aws
Google Cloud

AI GRAVITY PROBLEM

Hyperscale growth is driven by the race to build AI capacity. Providers focus on GPU-intensive, AI-optimized data centers, shifting traditional workloads into costlier tiers due to training with large GPU hours, inference data egress, and expanding data storage for datasets and models.

As AI demand drives new consumption models, pricing predictability declines. AI workloads need capabilities that don't fit traditional frameworks, especially for inferencing. As enterprises consolidate more data, moving or rebalancing workloads becomes increasingly complex and costly. This creates tension between scaling innovation and operational control, causing executives to reassess where true value lies.

RAW INPUT DATA
Data is cleaned/ normalized and aligned with AI model requirements

TRAINED MODEL

DATA PROCESSING
Algorithmic processing with GPU/CPU

OUTPUT
Output fed into the model for usage in predictions, decisions, or generalized context

Inferencing is crucial because it is the process by which a trained model takes input data and produces an output. It's heavily used in chatbots, which utilize natural language processing to understand and respond to user queries.

Handling unseen data and making predictions drive inferencing success, but doing so demands high throughput and low-latency interfaces which are costly and can lead to unexpected expenses for businesses. Managing compute, egress, and networking costs becomes complex as data moves into increasingly expensive cloud environments. Yet, this is crucial for model customization and usefulness.
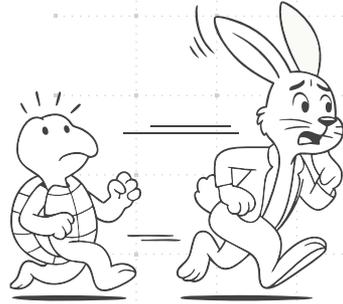
Introduction

Challenges

Transformation

Solutions

Citations

The Rise of Shadow AI Amid Frustration and Shortages

PACE OVER PROCESS

When internal cloud operations cannot keep up, business units often bypass IT to get what they need, which is known as the "pressure to get it done." Recently, "Shadow IT" has evolved into "Shadow AI"—which refers to employees using unapproved GenAI tools in defiance of company policy and security guidelines. The consequences can be severe; not only is there a significant risk of confidential data leaks, but also compliance violations and security weaknesses.

Data Exposure and Compliance Blind Spots

Staff experimenting with AI models on public platforms risk sensitive data leaving secure environments. Lack of governance can lead to external access to confidential information such as code, customer data, or intellectual property (IP), risking compliance and sovereignty. Frameworks such as the Digital Operational Resilience Act (DORA), the General Data Protection Regulation (GDPR), and the European Union Artificial Intelligence Act (EU AI Act) make this exposure costly and traceable.

Unvalidated Outputs and Model Risks

Teams may rely on AI outputs without verifying accuracy, explainability, or bias controls. Hyperscale models produce quick results but lack transparency in training data and decision logic, undermining auditability and risking errors, misinformation, or unfair outcomes in business processes.

Cost and Duplication of Effort

Shadow AI results in redundant spending due to multiple teams using different subscriptions, leading to fragmented budgets, resource waste, and hindered negotiations and risk management.

Technical Debt and the Innovation Squeeze

Organizations anticipated that the cloud would simplify commercial and technology usage, but it introduced new complexities, hindering innovation and increasing technical debt. As companies manage AI and develop strategies for public, private, and edge environments, technical debt remains a significant yet often overlooked obstacle to modernization, having a negative impact on numerous aspects of the business.

Legacy Entanglement Across Environments

Decades of accumulated code, custom integrations, and legacy dependencies don't vanish when workloads shift to the cloud—they build up. Moving old applications without re-architecting them just transfers the issues. The result: hybrid environments that are more difficult to manage, more fragile, and resistant to automation.
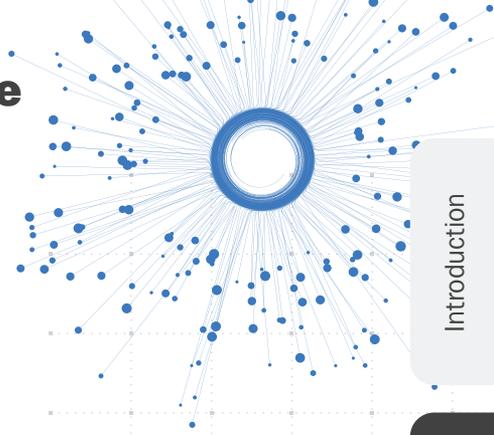
Cloud Fragmentation

The usage of clouds and tools has evolved faster than governance can keep up. Deploying new services to address new initiatives allows temporary increases in productivity, but without retiring old systems, it creates overlapping systems, duplicate data, and rising costs. Instead of innovating, IT spends more time maintaining outdated infrastructure.

Innovation Held Hostage by Maintenance

Budgets and talent are increasingly used to "keep the lights on." Time spent handling outdated workloads takes away from advancing automation, AI, and customer innovation. Without a clear plan to reduce technical debt, cloud adoption becomes a treadmill—movement without progress.
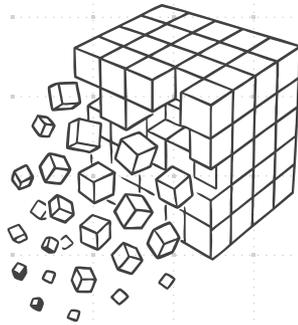
TRANSFORMATION

The question isn't whether to change—it's how to turn disruption into acceleration.

KEY SIGNALS OF CLOUD MATURITY

	Market Dynamics and Competition	Proactive Operational and Strategic Focus	Technology Advancement as a Change Catalyst
SIGNAL	Market Dynamics and Competition	Proactive Operational and Strategic Focus	Technology Advancement as a Change Catalyst
TRANSFORMATION	Stabilization is crucial for mitigating future disruption	Tools and solutions must focus on optimization	Organizations must stay ahead of competitors
SUPPORTIVE SIGNALS	✅ Slower cloud growth ✅ Market consolidation ✅ Fewer new entrants into the market ✅ Shift from innovation to GTM	✅ FinOps cloud optimization ✅ Automation of infrastructure ✅ Governance and security ✅ Cloud-native architectures ✅ Focus on business outcomes	✅ Integration of AI/ML into cloud ✅ Hybrid and multi-cloud as standard ✅ Tooling for hybrid and multi-cloud management
DISRUPTIVE SIGNALS	❌ Explosive growth ❌ Crowded market ❌ Focus on innovation, not monetization	❌ Spiraling cloud costs ❌ Shadow IT/AI ❌ Legacy and technical debt prevails	❌ Skills and resources in innovative tech ❌ Siloed cloud ❌ Tool sprawl ❌ Mounting technical debt

FRAME DISRUPTION AS A CHANCE TO RESET

Drivers and Outcomes

Cloud transformation now emphasizes removing ambiguity, optimizing, orchestrating, and modernizing across all environments. The focus has shifted from location to data-driven commercial optimization, leveraging data, applications, and infrastructure to achieve faster results, increased resilience, and measurable value.

Examine the focused questions:

? Which workloads belong in the public cloud versus private or edge clouds?

? How can I reduce risks in my ecosystem?

? How do we build portability and choice into future architectures?

? Where does AI influence our compute strategy beyond analytics?

The correct answers vary by organization, but the wrong answer is to do nothing.

TRANSFORMATION DRIVER	WHAT IT LOOKS LIKE	BUSINESS OUTCOMES
Cost Governance (FinOps)	Providing unified visibility across clouds, enforcing real-time accountability, and automating infrastructure pipelines	Up to 40% cost savings; 10–20% savings from FinOps as code[8, 9]
Commercial Optimization (Avoidance of Broadcom type issues)	Simulating extreme usage, audit, and acquisition to stress test models, build exit options, and reassess partner ecosystems	Proactive risk identification, continuity validation, and optimized licensing for strategic planning
Application Modernization	Re-architecting legacy applications with containers and APIs	Faster release cycles and improved scalability
Operational Automation	Streamlining workflows and auto-remediation of issues	IT efficiency gains without additional headcount
Hybrid Orchestration	Dynamically balancing workloads across cloud(s) and edge sites	Better performance and compliance control. Hybrid environments give organizations greater flexibility to balance security requirements with performance needs.
AI Integration	Embedding AI in infrastructure and operations, not just business apps	Predictive operations and smarter capacity planning

Where the Opportunity Lies

Cloud transformation evolves to a future state balancing hyperscale cloud efficiency, private control, and edge agility. Organizations must tackle licensing risk, hyperscale dependence, and technical debt to turn friction into flexibility and build adaptable architectures.

STATUS QUO

Rethinking the Financial Architecture

The market consolidation of cloud vendors in 2025 isn't new; it will likely continue as long as tech markets remain volatile, driven by the overvaluation of AI and quantum companies and debt-based trading. Many providers shifted from perpetual to subscription licensing, leaving businesses with:

LICENSE VOLATILITY
Multiple increases in contract costs, either retired or repackaged under new subscription terms.

LESS FLEXIBILITY
New license requirements, bundled offerings (and EOS of existing offerings), hardware support changes, and core threshold updates hinder many businesses from managing and scaling costs effectively, while tech vendors optimize and cut operational costs, inflating their valuations.

RISKY VENDOR DEPENDENCY
Many tech companies excel but face debt risks. While high market valuations based on future revenue predictions seem fine during strong markets, a downturn could lead to acquisition or closure. Relying heavily on one core supplier is risky.

Steps to take:

01 Perform licensing stress tests to project future exposure under new terms.

License Stress Test

TECHNOLOGY ROI

02 Reassess partner ecosystems to ensure service continuity and maintain a balanced multi-vendor approach.

Exit Contingency

Reassess Ecosystem

03 Design architectures with exit options to preserve leverage.

Business leaders now share responsibility for technology ROI. In an environment where costs can unexpectedly increase, commercial agility serves as a form of resilience. A successful organization can quickly adapt to changes in pricing, licensing, or vendor models, rather than cutting costs.

SOLUTIONS

Vendor Neutrality

Vendor-neutral, business-driven strategies prioritize flexibility over brand loyalty or short-term prices in a disrupted, consolidating market. Start with clear goals such as security, compliance, performance, agility, or cost control; then evaluate solutions objectively using consistent criteria.

01
BUDGET PREDICTABILITY

FINANCIAL
What's the total cost of ownership over three to five years, including renewals and exit costs?

02
REDUCED COMPLEXITY

OPERATIONS
How easily can we integrate, automate, and scale?

03
RISK MITIGATION

REGULATORY
Can the solution meet both sovereignty and compliance needs now and as laws continue to evolve?

04
LONG-TERM AGILITY

STRATEGIC
Does this choice align with our talent strategy, innovation roadmap, and data architecture?

Organizations should include choice mandates to ensure flexibility and avoid lock-in, using open standards, containers, and modular designs as a competitive advantage. Make it measurable with KPIs such as cost, deployment time, utilization, and compliance, and regularly reassess them. Treat transformation as ongoing, adapting to the rapidly changing cloud market and evolving innovation cycles.

Cloud Repatriation and Rebalancing

After a decade of a "cloud-first" approach, many businesses have come to realize that not all workloads fit hyperscale environments, especially following increasingly serious hyperscale outages. Repatriating workloads to private, colocation, or edge setups isn't a strategy change but fine-tuning to balance cost, control, and performance.
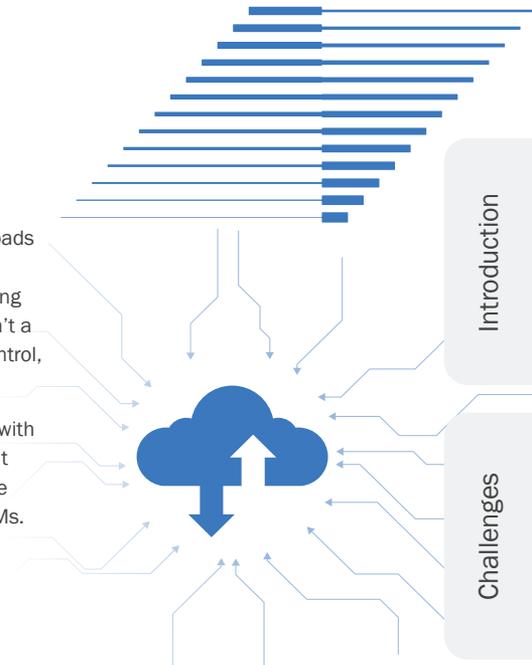
The process can be complex and costly, especially with large data egress and differing architectures. Still, it is faster when cloud providers and customers share platforms such as Kubernetes, unlike traditional VMs. A more straightforward approach is rebalancing— optimizing cloud deployment for cost, security, and performance not met by public cloud.

Key Considerations for Cloud vs. Local Workloads

AREA	CONSIDERATION
Cost	• Cloud reduces initial capital costs but adds complex, unpredictable operating expenses like storage, bandwidth, and per-use fees, which can exceed budgets. • Cost efficiency varies based on provider, architecture, and usage.
Security	• Public cloud isn't necessarily less secure, but misconfigurations and compliance gaps are common. • Physical control and straightforward auditing of on-premises or private clouds can be crucial for sensitive workloads.
Availability	• Outages are rare but can last days and affect thousands; recovery depends on the provider, which is out of organizational control. • Local workloads give organizations direct control and can be combined with cloud deployments for better redundancy and resilience.
Skills	• Cloud operations need expertise in architecture, automation, and monitoring. • Skills gaps raise risk, cost, and downtime, often leading to bringing services back in house or to managed environments.

Remember that moving workloads involves costs, so compare initial expenses with long-term benefits. The costs of change and work effort during assessment and migration, the expense of transitioning operational teams familiar with hyperscale tools to adopt new methods for a smoother transition, and the cost of running on-premises infrastructure can increase a company's carbon footprint unless it uses renewable energy or designs efficient data centers.

Repatriating workloads is most effective when guided by reliable cost analysis and workload baselines, not reactionary guesswork.

What's Driving the Swing Back

Whether you are looking at moving workloads from one hypervisor to another, rebalancing, or repatriating, **this is a correction to an issue**. Cost pressure, over-dependence, governance issues, shifting performance, and regulations all impact strategic choices for cloud endpoints.

Examine the signals for change. And consider the actions to take.

	FINANCIAL	OPERATIONAL	REGULATORY	ARCHITECTURAL	STRATEGIC
QUESTION	Are cloud costs stable, predictable, and justified by performance?	Do your teams have visibility and control over performance, availability, and configuration?	Are you meeting all compliance and data residency requirements?	Are workloads portable and cloud-native, or tightly coupled and dependent on vendor tools?	Does your current environment align with your innovation and AI roadmap?
SIGNAL	Persistent overspend and/or high egress charges	Lack of optimization or reliance on provider SLAs	Audit findings, new laws, or regulator pressure cannot be met	Vendor lock-in or limited migration options to other platforms, lack of control	Delays in AI enablement, stuck in AI POC paralysis or rising dependency risks
ACTION	Consider rebalancing to private or managed infrastructure for cost stability.	Redistribute workloads closer to users or edge environments for performance gains.	Repatriate or shift critical workloads where governance and audit readiness are stronger for sovereignty/ security.	Evaluate hypervisor diversification or multi-cloud models to regain leverage and choice in platforms.	Modernize or migrate workloads to environments that support scalable innovation.

The 4A Cloud Decision Framework

As noted by best practices such as the Gartner "cloud smart" approach, transformation is a process, not a one-off project. Exact Market recommends examining the following areas in detail, aligned with the **4A framework**, to justify rebalancing decisions and optimization.

01
Workload fit
Right environment for each use case

02
Financial clarity
Total cost across lifecycle

03
Operational agility
Portability and resilience

04
Governance discipline
Policy-driven placement and monitoring

Assess
- Inventory workloads, applications, costs, and dependencies across hypervisors, clouds, and edges.
- Build a clear baseline of where workloads run and why.
- Identify gaps in cost, compliance, and performance.

Align
- Align technology placement with business priorities.
- Map workloads to objectives: innovation, control, compliance, or cost efficiency.

Act
- Decide: stay, shift, rebalance, or repatriate.
- Apply financial, operational, and regulatory criteria to each workload. Document rationale and ownership.

Adapt
- Review quarterly and refine.
- Establish governance to continuously measure the alignment of cost, performance, and compliance.

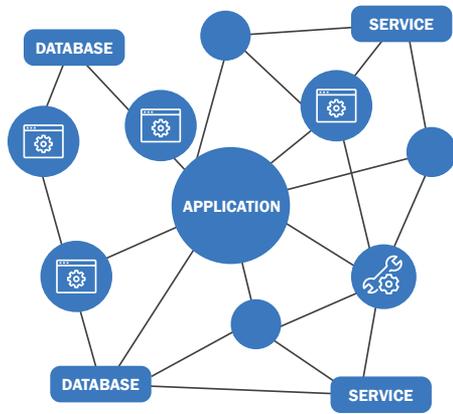# Understanding Application Dependencies

Before assigning workloads, understand system interactions, connections, data flows, APIs, and supporting services. Without this, migrations risk outages, slowdowns, or security issues.

Many organizations don't realize how their applications interconnect; a workload relies on various support services including authentication, data management, analytics, monitoring, and third-party tools. Ignorance of these links is a security risk, often only noticed during incidents. When one component moves, all connections must be considered.

Dependency mapping helps you see everything clearly, making mitigation and migration plans more effective and preventing unexpected dependencies from causing delays or extra costs.

01
Inventory applications and services.

Create a unified catalog of applications with ownership, dependencies, and lifecycle stage.

02
Map data, track flow between systems, assess criticality, and rank dependencies by impact if disrupted.

04
Feed insights into the 4A decision framework.

Use dependency data for workload placement.

03
Visualize and simulate migration, rebalancing, or repatriation effects using modeling tools.
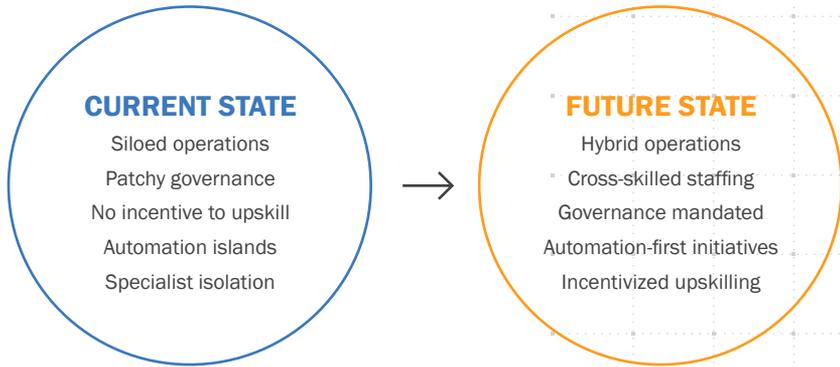
ACCURATE WORKLOAD PLACEMENT

Addressing Knowledge Gaps and Skills

The modern IT landscape demands skills in hybrid architectures, multi-cloud, and AI automation, but these skills are scarce. As cloud strategies evolve, the bottleneck is talent, not technology. Bridging knowledge gaps is a major leadership challenge.
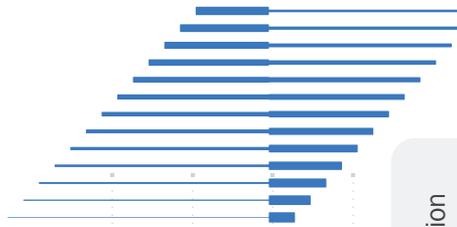
Reflect on how new technologies are prioritized and their impact on your business. A key factor is the increasing importance of technologies, such as AI agents and post-quantum cryptography. Developing skills and resources in these crucial areas will soon become essential.

CURRENT STATE

Siloed operations

Patchy governance

No incentive to upskill

Automation islands

Specialist isolation

FUTURE STATE

Hybrid operations

Cross-skilled staffing

Governance mandated

Automation-first initiatives

Incentivized upskilling

BUILDING RESILIENCE THROUGH CAPABILITIES

Cross-Train for Hybrid Skills	Promote lateral learning, such as cloud engineers learning edge and on-premises orchestration, or sysadmins learning automation and APIs. Hybrid fluency lessens dependence on single-role focus.
Implement FinOps and SecOps Practices	Redistribute cost and security responsibilities from isolated teams into shared accountability models. Integrate governance into daily workflows to ensure financial and compliance issues are managed as integral parts of regular operations, not isolated procedures.
Use Partner Ecosystems Wisely	Boost capability development without dependency via partnerships. Clearly specify knowledge transfer in contracts and monitor internal skill growth as success metrics.
Utilize Automation and AI for Scalability	Offset staffing shortages with innovative tools such as AI to identify anomalies, manage incidents, automate tasks, and create guidelines for less experienced staff.
Promote Continuous Learning	Make training central to transformation by promoting certifications, peer mentoring, and rotational roles in cloud, security, and data teams.

Cloud Options

Businesses should evaluate their cloud usage, as there are numerous options in cloud technologies, and each current setup should be reviewed in light of future requirements.

💡 BUSINESS AND STRATEGIC NEEDS

Business objectives and goals

Budget and cost (pay-as-you-go, hidden charges, discounts)

Scalability and elasticity requirements

Team expertise and development speed

Vendor lock-in concerns and exit planning

⚙️ TECHNICAL REQUIREMENTS

Required service model (IaaS, PaaS, SaaS)

Performance and reliability (SLA, latency, uptime)

Integration with existing tools/ecosystem (Microsoft, Google, etc.)

Data migration support

Deployment model (public, private, hybrid)

CHOOSING THE RIGHT CLOUD SOLUTION

🔍 VENDOR EVALUATION

Provider reputation and company profile

Quality of customer support and community resources

Service catalog breadth (e.g., major providers including AWS/Azure/Google Cloud)

🛡️ SECURITY AND COMPLIANCE

Data security measures (encryption, firewalls, IAM)

Regulatory compliance and certifications (GDPR, HIPAA, ISO)

Data residency/storage location requirements

Disaster recovery and backup processes

Identity and access management (IAM)

Choose a cloud solution—hyperscale, colocation, or endpoint—based on your business needs. Align IT goals with strategic priorities, the **"what,"** and set guardrails for risk, compliance, and budgets. This defines clear business objectives and the success of KPIs.

Technical requirements support your strategy and clarify **"how"** to implement it.

Next, after understanding which service models you need (IaaS, PaaS, SaaS, etc.), assess risk profiles and tolerances; not everything will fit neatly into these models due to concerns over data privacy, compliance, and criticality.
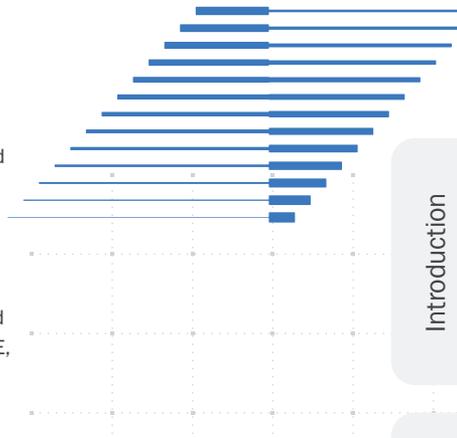
Finally, examine potential vendors or suppliers, understand the commercial implications, and evaluate their support, roadmaps, historical behaviors, leadership change impacts, and financial stability, as these are vital to your business success.

After creating a shortlist, test your model with real workloads, identify skill gaps, explore managed service support, and refine your strategy.

Hypervisor Options

There is no one-size-fits-all solution for on-premises cloud options. Server virtualization uses complex technologies that enhance utilization, portability, automation, and availability. The right hypervisor depends on volume, skills, use cases, and hardware support.

The most dominant market players (excluding those used by hyperscale) are VMware vSphere, Nutanix, Proxmox VE, Hyper-V, XCP-ng (Xen), and OpenStack, each with its own place and capabilities.

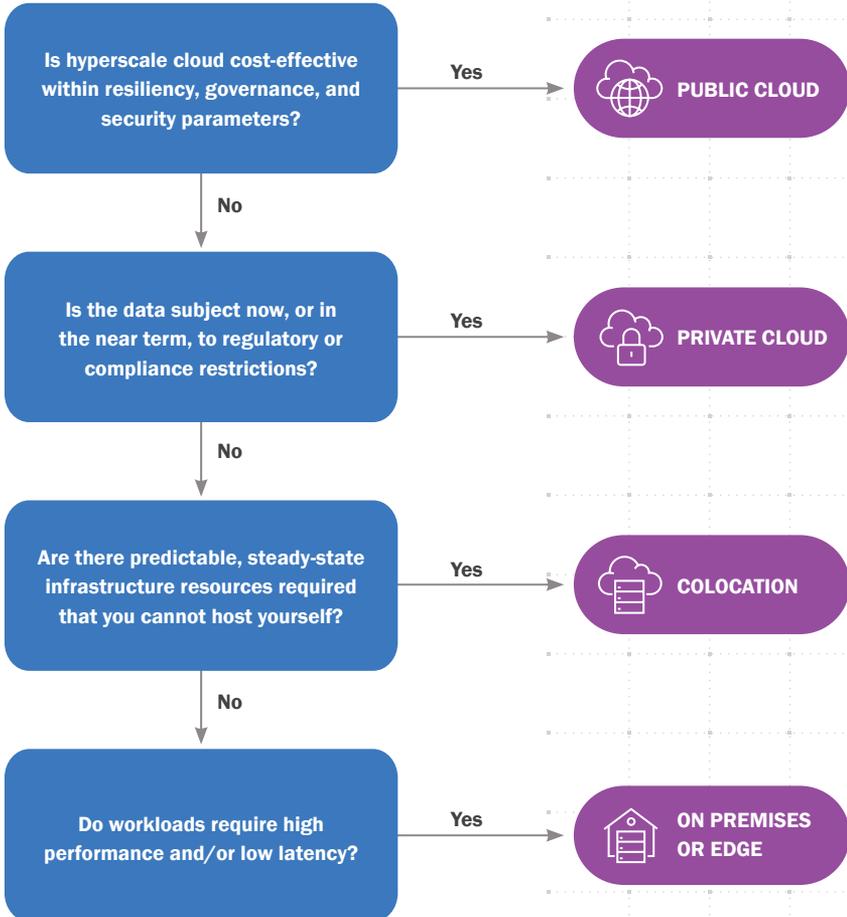| FEATURE	VMWARE	NUTANIX	HYPER-V	OPENSTACK
Architecture	Hardware virtualization with storage HCI (vSAN)	HCI appliance for compute and storage	Windows Server hypervisor	Open-source platform for infrastructure
Management	vCenter server (mature and powerful)	Prism (unified and intuitive)	Windows Admin Center/SCVMM (complex and powerful)	Multiple components and APIs (complex and DIY)
Licensing	Subscription/core/term (typically most expensive)	Node based (including Acropolis hypervisor)	Included with Windows Server (hence popular)	Open-source/free but requires investment to get working
Ecosystem	Third-party hardware/software compatibility, complex	Smaller ecosystem, tight integration, less choice	Integrated with Microsoft and Azure products	Fragmented ecosystem, lots of options
Scalability	Highly scalable and proven	Linear scale out adding nodes	Scalable within Windows environment	Highly scalable, horizontal for very large clouds
Ideal Scenario	Large enterprise	Large or medium enterprise, focus on simplicity	Microsoft businesses seeking cost effective hypervisor products	Large organizations, highly skilled staff and support

High-Level Decision Tree

In the 4A model, we analyze the decision-making process to identify strategic factors that influence how workload transformation is approached. While understanding portfolio planning and re-evaluation is important, it does not make sense at the workload level. Conversely, the decision tree prompts consideration of the conditions that affect workload placement.

Use the 4A model to design the process: determine **who** makes the decision, **what** data informs it, and **when** it should be reviewed. The next stage uses the decision tree to guide the endpoint choice, translating governance principles into specific placement outcomes.

Is hyperscale cloud cost-effective within resiliency, governance, and security parameters? → Yes → **PUBLIC CLOUD**

No ↓

Is the data subject now, or in the near term, to regulatory or compliance restrictions? → Yes → **PRIVATE CLOUD**

No ↓

Are there predictable, steady-state infrastructure resources required that you cannot host yourself? → Yes → **COLOCATION**

No ↓

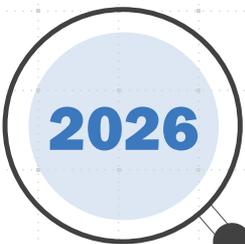**Do workloads require high performance and/or low latency?** → Yes → **ON PREMISES OR EDGE**

Looking Forward to Infrastructure 2.0

AI CONSIDERATIONS
Today, Next Year, and Beyond

AI is pervasive in cloud tech, reshaping decisions in compute economics, data architecture, and workload strategy. The question isn't if AI changes your environment, but how quickly you can adapt.

2026

TODAY — Optimizing GPU and CPU Placement

Today, organizations face the challenge of balancing capacity and costs. AI is suited for high-performance platforms with powerful GPUs, but not all applications need such power. Companies must decide when to use GPUs versus CPUs to meet needs without overspending. GPUs are now seen as valuable, costly assets crucial for goals, given market supply and demand.

2026 — Edge Orchestration and Inference Locality

As AI models grow, inference happens closer to data sources. Organizations distribute AI across cloud, edge, and devices to cut delays, protect data, and respond faster. This mix needs new orchestration layers—intelligent systems that allocate workloads between centralized and local computing based on rules and context.

2027 AND BEYOND — AI-Native Infrastructure

Looking ahead, AI will become core to our infrastructure, transforming cloud architectures with self-adapting solutions for better performance, energy efficiency, and compliance. Infrastructure will be AI-driven, seamless, and integrated.

Summary

Cloud transformation isn't about chasing new solutions but about regaining control, understanding choices, and optimizing for the future. Amid licensing changes, hyperscale providers adjusting prices, and skills gaps, the real advantage lies in intentional decision-making.

Disruption from the status quo offers a chance to rebalance, redefine controls, and realign costs, compliance, and capabilities through a vendor-neutral, data-driven, adaptable framework. The future isn't about fully committing to one cloud but about making informed, clear choices that evolve with business needs.

Define strategic transformation goals with the **4A model**

Utilize the workload **decision tree** to understand workload placement

Understand your organization's capabilities and future-proof your strategy

Encourage learning and development to cross-skill and maintain staff resources

Introduction

Challenges

Transformation

Solutions

Citations

CITATIONS

CITATIONS

1. Gartner, Gartner's Insights on the Future of Cloud Computing: 2025 Report, Jan 2025

2. United Nations, Data Protection and Privacy Legislation Worldwide, July 2025

3. ECCO, 2nd ECCO Report on Broadcom Appendix, May 2025

4. Opscompass, VMware Licensing in 2025, Apr 22, 2025

5. Network Computing, Customers Frustrated with VMware after Broadcom Acquisition, Apr 2, 2025

6. Forrester, Predictions 2025: Accelerated Demand For AI-Powered Infrastructure And Operations, Oct 22, 2024

7. Goldman Sachs, AI To Drive 165% Increase in Data Center Power Demand by 2030, Feb 4, 2025

8. Deloitte, Cloud Gets Lean: 'FinOps' Makes Every Dollar Work Harder, Nov 19, 2024

9. McKinsey, Everything Is Better as Code: Using FinOps to Manage Cloud Costs, Feb 3, 2025

ABOUT

exactmarket ™

Exact Market

is here to help your organization understand and capitalize on technology disruptions.

We bring together market insights, content strategies, and digital execution to help enterprises, technology providers, and independent software vendors share complex transformation launches and stories clearly and compellingly.

Our Elephant in the Room series captures the meaningful conversations leaders are already having today—about cloud, cybersecurity, AI, and infrastructure—and transforms them into practical frameworks for making informed decisions.

Whether you are refining your cloud strategies, launching new products, or preparing for the next wave of innovation, Exact Market's specialized marketing resources can help align your business, technology, and brand.

> **WE DON'T JUST EXPLAIN TRANSFORMATION**
> **WE GUIDE YOU IN SHAPING IT, OWNING IT, AND MOVING IT FORWARD.**

Find out more @ www.exactmarket.com