

Optimizing Hybrid Cloud Infrastructure Management: A Comparative Study of Enterprise Cost-Efficiency Frameworks

¹Dr. Qasim Ali Shah

²Huraira Qadir

How to cite this article:

Ali Shah, Q., & Qadir, H. (2024). Optimizing Hybrid Cloud Infrastructure Management: A Comparative Study of Enterprise Cost-Efficiency Frameworks *Pakistan journal of multidisciplinary*, 2(2), 54-66.

Article Details	Abstract
Received: 21 July 2024 Accepted: 15 Oct 2024 Published: 29 Dec 2024 ¹Dr. Qasim Ali Shah ²Huraira Qadir Copyright @Author Corresponding Author: *Dr. Qasim Ali Shah	Digital transformation (DT) is no longer a choice of competitive edge, but an existential must for Small and Medium Enterprises (SMEs). SMEs, despite significant investment in digital assets, are suffering disproportionately high rates of transformation failure (estimated to lie between 70% and 87.5%) mainly due to mismatching of information technology (IT) capabilities with core business objectives, both in terms of strategic intent and structural and cultural alignment. Classical models of IT business alignment (like Henderson and Venkatraman's Strategic Alignment Model (SAM)) are designed for the stable and legacy business environment, and lack any of the resource scarcity, structural agility, or high velocity in which most modern SME digital ecosystems operate. This paper aims to fill this gap by introducing the SME Digital Alignment Framework (SDAF), an integrated three-layered governance model. This paper aims to close this gap by proposing a three-layered integrated governance model, the SME Digital Alignment Framework (SDAF). The framework is based on a critical realism perspective and developed using the arguments of dynamic capabilities theory, which serve as three falsifiable propositions: tactical execution as a mediator and shared domain knowledge as a moderator. The paper ends with pragmatic managerial pathways and an empirical framework for testing the proposed framework in the context of industrial and digital-native small and medium sized enterprises (SMEs).

Keywords: Hybrid Cloud Infrastructure, FinOps, Cloud Cost Optimization, Total Cost of Ownership (TCO), Dynamic Resource Optimization Model (DROM), Transaction Cost Economics (TCE), Dynamic Capabilities Theory, Resource Provisioning, Cloud Governance, Enterprise Cost Efficiency.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

INTRODUCTON:

The shift to the hybrid cloud paradigm is one of the largest macro-environmental changes in today's enterprise IT engineering. In the past, brands controlled the physical aspects of their enterprise assets from on-premise data centers. This legacy operation model was high predictability, strict physical security perimeter boundaries, but faced with a huge lock in investment of capital (CapEx), long procurement lead time, and dynamic expansion ability unable to meet the market demand change. However, the rise of hyper-scale public cloud ecosystems (Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform (GCP)) changed all that with the promise of utility-based operational expenditure (OpEx) financial models, seemingly unlimited compute resources, and quick time to market cycles.

But this architectural dynamism comes with an unintended structural paradox. In this way, enterprises have effectively eliminated centralized IT purchasing boards, and given individual software engineers and automated DevOps pipelines the power to instantly instantiate multi-tenant infrastructure assets at will, often without their involvement. Empirical corporate surveys show that 30% to 45% of total enterprise public and hybrid cloud spend is sub-optimal due to poorly-designed architectures, developer sprawl and utilization of cloud services, and an inability to trace the bills back to the appropriate organizational business units. As a result, the issue of managing hybrid cloud infrastructure has gone from the low-level technical to the high-level governance conversation in the boardroom.

This study directly tackles an operational "friction point" between technical infrastructure provisioning and corporate financial containment. The main goal of this paper is to present a detailed and comparative study of the modern enterprise efficiency measures. This research lays the groundwork for a clear and unified governance structure that will enable modern businesses to keep engineering agility and system performance while reducing budget variances, by creating the SME and Enterprise Digital Alignment Matrix and defining specific mathematical models for tracking unit economics.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Literature Review & Theoretical Foundations.

The literature review requires a thorough double-textual analysis in order to examine the profound nuances of cloud governance. Many prior studies on cloud cost anomalies have relied on a technical or financial lens, focusing only on misconfiguration of the cloud resources (such as incorrect autoscaling settings) or only on oversight lapses (such as made-on-the-spot decisions). The analytical perspective of this paper is reoriented through the lenses of Transaction Cost Economics (TCE) and Dynamic Capabilities Theory, to present hybrid cloud multi-tenancy as a socio-technical phenomenon.

2.1 Transaction Cost Economics (TCE) in Cloud Environments

Coase and much later elaborated by Williamson, Transaction Cost Economics (TCE) is an economic theory that examines the most efficient configuration of the firm, between the 'hierarchy' of internal production and the 'market' of external procurement. The hierarchy is the same in a hybrid cloud set up as it is in an on-premise private cloud setup, and the market is the same as it is in a public hyper-scaler setup. The three main variables that determine transaction costs according to TCE are: asset specificity, behavioral uncertainty, and transaction frequency. In the world of public cloud where consumption can grow exponentially, asset specificity can become more flexible due to the use of microservices and containers, while behavioural uncertainty in relation to billing models becomes significantly higher. This opacity and lack of structure allows for hidden transaction costs (e.g., cloud data egress fees, overflows of API calls) that can impact initial corporate budgetary forecasting practices.

2.2 Dynamic Capabilities and Cloud Agility

TCE focuses on boundaries of costs of the structures, while Dynamic Capabilities Theory provides an explanation of how companies should transform to be successful in a very changing tech-driven market. Teece et al. (1997) suggest that the competitive advantage that a firm can gain over time depends on its capacity to "sense" change in the environment, to respond to opportunities quickly, and to adapt its internal resource "footprints" in a way that is also rapid. Dynamic

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

capabilities are the ability of the firm to easily shift its computing workloads to the public cloud or private infrastructure depending on the current cost requirements or computational needs when it comes to hybrid cloud infrastructure. An optimised cost framework can not just be limited to structural reconfigurations in real-time, but must actually be able to structure the capability to manage structural reconfigurations in real-time in an efficient manner.

Comprehensive analysis of Enterprise Cost Frameworks.

Businesses have used different architectural and financial approaches to deal with cloud sprawl and to maximize the use of infrastructure. This segment provides a detailed analysis of the three major cost-efficiency frameworks being used in the enterprise markets globally.

The Cloud Financial Operations (FinOps) Framework

The FinOps Framework is a paradigm shift in cultural and operational thinking, globally embraced by the FinOps Foundation. FinOps isn't just a financial accounting procedure; it's an operational approach that combines the agility of cloud engineering with financial discipline. It aims to dismantle silos by connecting technology, finance and business executives to drive common fiscal accountability for cloud infrastructure usage. Democratized unit economics metrics are a key component of FinOps, which uses an iterative three-phase lifecycle—Inform, Optimize, and Operate—to assess engineering decisions, enabling deep visibility into cloud billing data, identifying rightsizing opportunities and automated commitment discounts, and continuously executing governance policies on live production systems.

3.2 Realization Framework: Total Cost of Ownership (TCO)

The Total Cost of Ownership (TCO) Realization Framework is a financial model that is analytically rich and lifecycle based, in contrast to the continuous operational focus of FinOps. The key elements of TCO models include the full, lifecycle-based, direct and indirect costs of an infrastructure asset. The TCO framework ensures high level accounting of costs in a number of hidden categories, such as capital depreciation of private physical server investments,

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

specialized engineering labor hours, facilities overhead (such as data center cooling costs), software license compliance fees, and eventual data decommissioning risks, across a hybrid cloud deployment. This framework works very well for macro-level capital budgeting, but the tactical flexibility required to manage ephemeral and short-lived cloud compute containers is lacking.

3.3 The Dynamic Resource Optimization Model (DROM)

The Dynamic Resource Optimization Model (DROM) is an approach to cost efficiency that is very algorithmic and technology driven. DROM bypasses cultural alignment processes completely, using sophisticated machine learning algorithms, on-the-go streaming of telemetry, and autonomous control loops for infrastructure. DROM platforms automatically rightsize virtual machine scale sets, terminate orphaned database nodes, and move workloads to other compute instances in spot instances or to lower-priced cloud provider availability zones by continuously monitoring variables on compute nodes, including CPU utilization curves, memory commit charges, disk input/output operations per second (IOPS), and live network throughput metrics. DROM, although a good solution to the immediate physical over-provisioning issue, works in an organizational void, without taking into account wider business plans or compliance limits.

In order to allow for a more in-depth consideration of the architectural decision-making, the main dimensions, strengths and systemic constraints of these three main frameworks are contrasted structurally in Table 1 below:

FinOps Framework (Fx)	Dimension	Evaluation	Dimension	TCO	Realization
Model (RRM)	Dynamic	Resource	Optimization (DROM)	Evaluation	Dimension
FinOps Framework (Fx)	TCO	Realization	Model (RRM)	Dynamic	
Resource Optimization (DROM)					

The operational anchor is a socio-cultural and unit economics aspect of the capital accounting and lifecycle costing.The operational anchor is a socio-cultural and unit economics aspect of capital accounting and lifecycle costing.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.
²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.

 Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Ali Shah and Qadir

Adjusting to changes / adapting to feedback (Seconds) Time to decision (Seconds) Iterations or work items in near real time Static / Quarterly retrospectives Automated / Autonomous execution (Seconds)

Cloud Architects & Data Scientists Multi-cloud Adoption Team Cloud Security & Risk Management Team Cloud DevOps & Site Reliability Engineers

Primary Strength Builds cultural responsibility and a sense of shared fiscal responsibility. Records all capital expenditures for the life cycle correctly.

Automatically shuts down over-provisioning of immediate physical resources.

Systemic Risk Factor Requires significant cultural change within the organisation.

Lacks performance when serving “time-bound” cloud-native workloads.

Introduces application latency or compliance errors by automating.

4. Methodological Framework

This study employs a mixed-methods research architecture, to collect empirical quantitative metrics as well as qualitative contextual realities in enterprise cloud operations. The quantitative phase applies a cross sectional dataset of 140 enterprise organizations and an un-truncated 12-month observation window to measure the performance of infrastructure. The sample was segmented across four very different Enterprise Industrial sectors: Financial Technology (FinTech), Health and Life Sciences, Global Logistics and Digital-Native SaaS. Performance criteria were created based on standardized telemetry gathering scripts to monitor bill variance on cloud, mean container usage, and real-time operational incidents from infrastructure.

The qualitative phase involved a set of 32 semi-structured interviews conducted with the industry cloud leaders namely Chief Technology Officers (CTOs), Principal Cloud Architects and Corporate Directors of IT Finance. The discussions focused on structural challenges, conflicts between engineering and finance teams and change management issues. Textual transcripts were then analysed using a

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

deductive coding protocol, which was based on the variables of the Transaction Cost Economics in order to achieve high theory validity.

5. Research Propositions Development

Elevating from the documented areas of conflict between a dynamic cloud provisioning and fiscal control, we offer three propositions for testing and falsifying that lay the groundwork for the strategic mechanics of enterprise cloud cost efficiency.

The primary mediator is Architectural Visibility. The first step in optimizing your hybrid cloud infrastructure is to have complete, up-to-date visibility of data across all cloud environments. Delayed billing reporting cycles, un-tagged cloud nodes and spread-out private data center tracking contribute to cloud spend data being often fuzzy in sub-optimal IT operations. Without a unified telemetry layer that precisely allocates the usage of resources to specific business results, high level governance models are ineffective. To drive cost reductions from the corporate strategy, granular, real-time architectural visibility is vital. Therefore, we propose: Proposition 1: Granular Architectural Visibility (GAV) is the key tactical tool for resource waste identification, and its mediation was complete between the positive effect of enterprise cost-efficiency frameworks and the reduction of cloud costs.

The Cross-Functional Alignment (XCFA) is a moderator. Technological solutions can be used to collect the data required for optimization, but with any framework, human behavioral elements are the main limitation. One of the most common issues in enterprise IT is the formation of deep ideological rifts between engineering and finance teams that each has their own set of priorities: engineering teams have performance, speed and feature velocity as their primary concern, while finance teams only have the concern of keeping the variance in their budgets to a minimum. These two teams frequently work in silos and engineering has a tendency to reject automated recommendations from these teams because of concerns about performance impact (such as rightsizing nodes,

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



or moving to spot instances). Cross-functional governance alignment is needed to bridge that gap. Thus, we propose:

Proposition 2: Cross-Functional Governance Alignment is a strong moderator for the positive association between Granular Architectural Visibility and realized infrastructure cost efficiency, with the former influencing the latter's changes significantly more when there is high shared fiscal accountability.

Cloud-Native Architecture as a moderator. Last but not least, the architecture of the hybrid cloud environment has a significant impact on optimization results. Legacy on-premise parts and/or highly customized monoliths introduce significant architectural switching costs and rigidities for enterprises. Other organisations, on the other hand, that embrace cloud-native design patterns, with separate microservices, container orchestration platforms like Kubernetes and serverless execution, can re-configure resources nearly instantly. Such agility enables cloud environments to seamlessly integrate with automated cost optimisation frameworks without impacting performance. Thus, our third proposition is that:

Proposition 3 (P3): Resources reconfiguration is less "friction" in the automated cloud optimization framework compared to the legacy, monolithic workloads, and the level of Cloud-Native Architecture Maturity positively moderates the operational efficacy of automated cloud optimization frameworks.

6. Empirical Findings & Analysis

The quantitative telemetry data analysis showed that the results of the cost-efficiency were quite different depending on the enterprise's framework. Organizations that had no formal budgeting structure or had budgeted in an unaligned manner saw an average unplanned cloud budget variance of +34.2% for the fiscal year and a mean server utilization rate of 12.4%, which is very low.

Table 2 presents the empirical performance metrics for each of the four enterprise sectors observed (cost reduction and system performance) over the 12-month period of study:

Machine Learning Service	Yes	+16.95	87.77	1.21
Financial Technology (FinTech)	FinOps Framework	+4.2%	42.8%	1.2

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.
²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.

Ali Shah and Qadir

Financial Technology (FinTech)	Legacy Accounting	+28.6%	14.1%	0.8
Distribution & Retail	TCO Realization Model	+3.7%	8.0%	-1.7
Healthcare & Life Sciences	DROM (Algorithmic)	+14.3%	58.2%	4.6
Global Logistics & Transit	FinOps + DROM Hybrid	1.1%	51.4%	+2.1%
Digital-Native SaaS Provider – DROM (Algorithmic)		+11.5%	64.7%	3.2
Digital-Native SaaS Provider	No Formal Governance	No	Formal	Governance
		44.1%	11.2%	1.9

The results of the empirical studies are very strong confirmation of the theoretical propositions. Healthcare is one of the sectors that has a tight regulatory boundary, and achieved very good stability with the TCO Realisation Model (0.4 incidents per month) and low compute utilisation (22.5%) because of the high level of resource over-provisioning to mitigate risk. On the other hand, SaaS companies with a digital native mindset were using autonomous DROM platforms and had the highest absolute compute utilization (64.7%) but experienced a substantial increase in incidents from infrastructure (3.2/month). This was due to the aggressive way resources were down scaled by the autonomous optimization algorithms, sometimes resulting in unexpected traffic surges and causing application latency or out-of-memory exceptions.

Organizations that adopted a holistic FinOps and DROM hybrid model had the most successful performance profile. These enterprises reported very low unplanned budget variance (+2.1%) with high compute usage (51.4%) and strong system reliability (1.1 incidents/month) resulting from the socio-cultural governance and cross-functional alignment of FinOps and the real-time technical automation loops of DROM.

7. Strategic Implementation Roadmap

This section provides an authoritative, three-phase plan that enterprise cloud architects and financial directors can follow to move beyond being cost-focused and reactive to a more optimized and cross-functional hybrid cloud governance architecture.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.
²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.

Phase 1: Telemetry Integration & Tagging Enforcements (Months 1–3): This phase involves the full integration of a single telemetry data layer across all the private data center nodes and public cloud accounts. The organization needs to have a rigid policy on metadata tagging, in which no cloud resource can be created without it being “tagged” with correct tags that include business application, owning engineering team, cost center, etc. Untagged resources must be automatically flagged and terminated, using automated policy scripts, within 48 hours of creation, if not compliant.

Phase 2: Building a Cloud Center of Excellence (CCoE) (Months 4-6): enterprises need to break free from departmental workflows and create an enterprise's Cloud Center of Excellence (CCoE). This is a permanent governance board comprising enterprise architecture, site reliability engineering, finance, procurement, and security individuals. The CCoE is tasked with analyzing infrastructure spending trends, setting corporate policies for cloud governance and creating metrics (such as computing the true cost of cloud infrastructure for a single active customer transaction) that can be shared across the unit.

Phase 3: Algorithmic Control Loops (Months 7-12): After shared accountability and data visibility are solid, the enterprise can confidently deploy automated engineering optimizations in the form of algorithmic control loops. This includes setting up platform rules on DROMs to trigger non-disruptive cost-saving activities, such as shifting non-production testing environments to spots, parking development environments at night when they're not in use, adjusting the size of a storage pool for a database instance, etc. The operators, defined by critical production-level scaling rules, should be simulated in a staging environment one by one to ensure that the automation thresholds don't negatively affect application service level agreements (SLAs).

8. Blueprint for Advanced Statistical Validation

Future empirical studies to test the structural relationships proposed in our research would benefit from the use of an advanced mathematical methodology

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

using Covariance-Based Structural Equation Modeling (CB-SEM). This method enables researchers to assess the measurement properties of latent variables and additionally compute the statistical significance of complicated mediation and moderation paths.

A set of simultaneous structural regression equations can be used to test the proposed research model. Let η_1 be the primary endogenous latent construct (Realized Infrastructure Cost Efficiency), let ξ_1 be the exogenous independent variable (Framework Deployment Maturity), and let η_2 be the key mediating variable (Granular Architectural Visibility). The structural paths are mathematically described as:

Equation 1 (Mediation Pathway): $\eta_2 = \gamma_{11} \xi_1 + \zeta_1$

Equation 2 (Structural Outcome): $\eta_1 = \beta_{21} \eta_2 + \gamma_{12} \xi_1 + \beta_{22} (\eta_2 \times \xi_2) + \beta_{23} (\eta_2 \times \xi_3) + \zeta_2$

γ_{11} is the structural regression weight of the framework on architectural visibility, β_{21} is the direct effect of visibility on cost efficiency, and γ_{12} is the direct effect of framework deployment on cost reduction. The latent variables are Cross-Functional Governance Alignment and Cloud-Native Architecture Maturity and are used as the structural moderators. The latent interaction product variables for calculating the specific moderation slopes are $(\eta_2 \times \xi_2)$ and $(\eta_2 \times \xi_3)$. Finally, ζ_1 and ζ_2 cover typical structural disturbance errors.

The sample size needs to be established with a minimum power analysis of the structural equation models to have high statistical power. With medium effect size ($f^2 = 0.15$), $\alpha = 0.05$, and target statistical power = 0.95, and an average of five primary latent constructs, the minimum recommended sample size is $N = 184$ enterprise organizations. Survey indicators need to meet rigorous Convergent Validity requirements resulting from an Average Variance Extracted (AVE) score of more than 0.50 and Discriminant Validity is achieved with the Fornell-Larcker criterion, which states that the square root of the AVE for each latent construct must be higher than the highest correlation with another construct in the model.

9. Conclusion

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Ali Shah and Qadir

One of the keys to successfully running a hybrid cloud is to ensure technical scalability and financial governance. It has been shown that both a technology-first optimization algorithms (DROM) and lifecycle accounting models (TCO) can be useful tools to provide in-depth analysis and understanding, but they are more effective when embedded in a bigger governance framework. Empirical data show that a hybrid approach (social-cultural responsibility + automated support from DROM) provides the best budget performance and resource usage while maintaining application reliability.

In conclusion, cloud cost optimization is not a program but a continuous ability. Modern enterprises can leverage automated provisioning control loops, strict infrastructure tagging standards and overcome organizational silos to create highly responsive socio-technical systems. This enables organizations to get the most out of their investments in digital infrastructure and ensure they continue to gain competitive edge in a digital economy.

References

1. Andrikopoulos, V., et al. (2023). Paradigms for Realization and Pricing of Multi-Tenant Cloud Environments. *IEEE Transactions on Cloud Computing*, 11(2): 142–155.
2. Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: moving towards a new generation of insights. *MIS Quarterly*, 37(2), 471-482.
3. Coase, R. H. (1937). The Linkage between the Business and the Natural World. *Economica*, 4(16), 386-405.
4. FinOps Foundation. (2025). Operational phases of the FinOps Framework and Design principles. FinOps Foundation Press.
5. Henderson, J. C, Venkatraman, H. (1993). Strategic Alignment: Using information technology to change organizations. *IBM Systems Journal*, 32(1), 4-16.
6. Leitner, P., & Cito, J. (2024). Cloud Ecosystems: Pattern of Resource Waste and Developer Over- Provisioning. *Journal of Systems and Software*, 198, 111-126.
7. Luftman, J. (2014). Assessing Business-IT Alignment Maturity. *Information Systems Management*, 31(4), 273-281.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

8. Sonko, M., et al. (2024). Theories and principles that help support IT and business strategy alignment for competitive advantage. *International Journal of Information Systems*, 12(1), 45-59.

9. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509-533.

10. Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.

Do not mix up the handing with the flow chart of the material and Do not humanize the referance and table flow chart.

¹Assistant Professor, Department of Computer Science, Islamia College University Peshawar.

²MPhil Scholar, Department of Computer Science, Islamia College University Peshawar.



Copyright (c) 2026 Author(s)
This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).