INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

A Review of Cost Management Approaches in Cloud Computing

[1]Ankit Pandey,[1]Rounak Bhardwaj,[1]Keshab Mondal,[1]Ritesh Choudhry, [2,*]Pradip Ghanty

[1] Final Year Students of B. Tech in Computer Science and Engineering (Data Science), Kazi Nazrul University, Asansol, West Bengal, India.

[2]Assistant Professor & Coordinator, Department of Computer Science, Asansol Girls' College, Asansol, West Bengal, India

***Corresponding author**

DOI : https://doi.org/10.51583/IJLTEMAS.2025.1411000084

Received: 28 November 2025; Accepted: 03 December 2025; Published: 18 December 2025

ABSTRACT

Cloud computing has revolutionized the IT industry by offering scalable and on-demand resources under a pay-as-you-go model. However, the same flexibility that makes cloud computing attractive also introduces challenges in managing costs effectively. Cloud Cost Management (CCM) focuses on monitoring, analyzing and optimizing expenditure to achieve an ideal balance between performance and affordability. This paper reviews existing cost management strategies, techniques, and tools used across different cloud environments. It highlights traditional budgeting and monitoring methods, advanced optimization models using machine learning, and automation through FinOps practices. There view also identifies current challenges such as complex pricing structures, lack of cost visibility in multi-cloud environments, and unpredictable billing patterns. Finally, the paper discusses emerging trends and future research directions aimed at intelligent and sustainable cost optimization in cloud ecosystems.

INTRODUCTION

The rapid adoption of cloud computing [8] has fundamentally transformed how organizations deploy and manage IT infrastructure. Instead of maintaining costly on-premises systems, businesses increasingly rely on cloud service providers such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) to access computing resources on demand. While the pay-as-you-go pricing model offers significant flexibility and scalability advantages, it frequently results in unpredictable and escalating expenses when cloud resources are not managed efficiently.

As cloud environments grow in scale and complexity, Cloud Cost Management (CCM) has emerged as a critical focus area for both enterprises and researchers. Effective CCM ensures optimal allocation of cloud resources [11], thereby minimizing over-provisioning, reducing idle resource usage, and improving overall financial efficiency. CCM encompasses a combination of financial monitoring, operational governance, and technological optimization practices that aim to balance performance requirements with cost constraints.

Although major cloud platforms provide cost-monitoring dashboards and pricing calculators, these solutions often prove insufficient in handling complex cost dynamics in multi-cloud environments [3]. Differences in pricing models, billing metrics, and service abstractions across providers limit unified cost visibility and centralized governance. To address these limitations, researchers and practitioners have proposed a wide range of approaches, spanning static budgeting and rule-based monitoring techniques to dynamic resource scheduling mechanisms and AI-driven cost prediction models [9].

Despite the availability of diverse cost management strategies, existing studies are often fragmented, focusing on individual techniques without offering a comprehensive comparative perspective. This review paper aims to systematically analyze and synthesize the existing literature on cloud cost management, evaluate the effectiveness and limitations of current approaches, and highlight emerging trends that have the potential to

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

shape the next generation of cloud cost optimization. The ultimate objective is to provide a consolidated understanding of how cloud cost management strategies evolve to meet the growing demands of scalable, sustainable, and intelligent cloud ecosystems.

REVIEW METHODOLOGY

This review follows a structured and systematic literature analysis to examine existing approaches in cloud cost management. Relevant research articles were collected from widely recognized digital libraries and academic repositories, including IEEE Xplore, SpringerLink, ACM Digital Library, ScienceDirect, and arXiv.

The selection process focused on peer-reviewed journal articles, conference papers, and high-quality preprints published between 2010 and 2025, emphasizing studies related to cloud cost optimization, FinOps practices, AI/ML-based cost prediction, dynamic resource provisioning, and multi-cloud cost management. In addition to recent studies, a limited number of foundational works were included to establish baseline concepts in cloud pricing and resource allocation.

Papers were selected based on the following inclusion criteria:

i) Relevance to cloud cost management or optimization,

ii) Discussion of financial, operational, or algorithmic cost-control mechanisms, and

iii) Conceptual clarity or experimental validation.

Studies were excluded if they focused solely on performance optimization without explicit consideration of cost, lacked sufficient technical or analytical depth, or were unrelated to cloud-based environments.

The selected literature was systematically analyzed and classified into three major categories-static approaches, dynamic approaches, and AI/ML-based optimization techniques-based on their adaptability, automation level, and decision-making capability. This classification provides a coherent analytical framework to compare existing methods, identify limitations, and highlight emerging research directions in cloud cost management.

Cloud Cost Management Overview

Cloud Cost Management (CCM) refers to the set of strategies, tools, and practices used to monitor, analyze, and optimize the financial expenditure associated with cloud resources. In The pay-as-you-go model adopted by most cloud service providers, organizations are charged based on the resources they consume such as computing power, storage capacity, network bandwidth, and additional services. Although this model offers scalability and flexibility, it also introduces complexity in predicting, controlling, and optimizing costs, especially in multi-cloud or hybrid environments.

Effective CCM involves several key objectives

Cost Visibility - providing detailed insight into where and how money is being spent across cloud resources and services [6].

Cost Optimization-identifying underutilized or idle resources and adjusting usage to minimize waste.

Forecasting and Budgeting [14] –using analytics and predictive models to estimate future costs and plan budgets accordingly.

Governance and Accountability-ensuring that each team, projector department is responsible for its cloud expenditure.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

The major cost drivers in cloud computing include

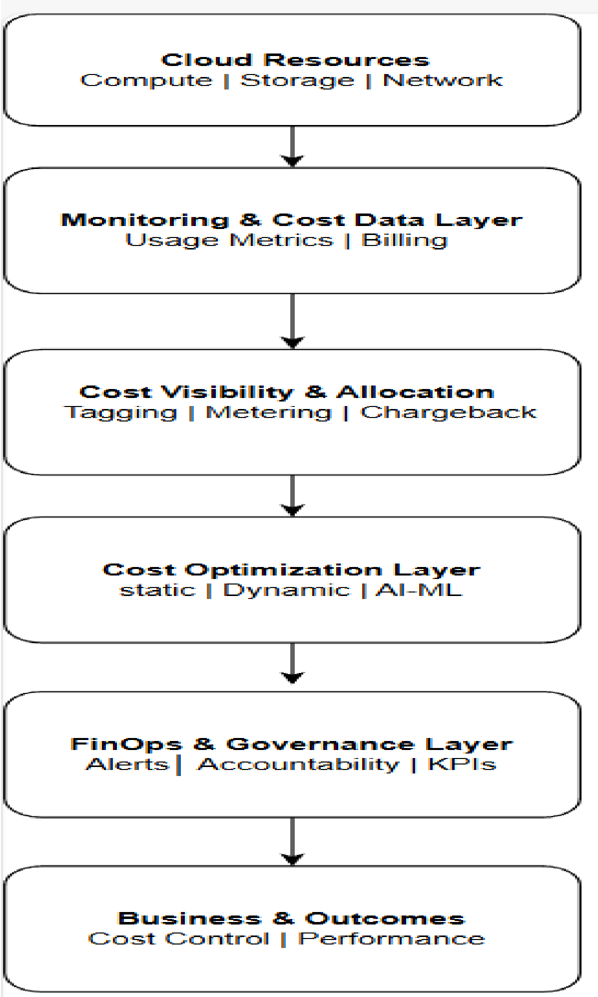

Figure 1: Conceptual Cloud Cost Management framework

Computer resources: Virtual machines (VMs), containers and serverless functions that run applications.

Storage: Costs associated with persistent storage, database services, and data backup.

Data transfer: Network usage costs, particularly dataegress between regions or providers.

Licensing and third-party services: Software subscriptions, APIs, and managed services integrated into cloud operations.

CCM frameworks often employ automation to control costs in real time through features such as auto-scaling [1], instance scheduling, and rightsizing. Additionally, practices like tagging (assigning meta data to resources) and FinOps (Financial Operations) help organizations align financial goals with engineering decisions.

In essence, Cloud Cost Management is not just about reducing expenses; it's about achieving an optimal balance between performance, scalability, and financial efficiency. As organizations increasingly adopt multi-cloud and hybrid models, CCM plays a critical role in ensuring operational sustainability and strategic decision-making.

Conceptual Framework for Cloud Cost Management

Figure 1 presents a conceptual Cloud Cost Management (CCM) framework that integrates technical cost optimization mechanisms with financial governance practices. The framework illustrates the flow from cloud resource consumption to monitoring, cost visibility, and optimization. Static, dynamic, and AI/ML-based techniques operate at the optimization layer to control expenditure, while FinOps practices enforce accountability, budgeting, and policy-driven governance. This layered structure highlights that effective cloud

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

cost management requires coordinated interaction between technical automation and organizational governance rather than isolated cost reduction strategies.

Existing Approaches and Techniques

CloudCostManagementhasevolvedthroughavarietyofapproaches,rangingfrombasiccost tracking methods to advanced AI-driven optimization frameworks. These techniques can broadly be categorized into static, dynamic, and intelligent (AI/ML-based) approaches. The approaches are described in subsections. In Table 1 Comparison of Cloud Cost Management Approaches are depicted.

Static Approaches

Static approaches to cloud cost management represent the most traditional layer of cost-control mechanisms. These approaches rely on predefined rules, fixed budgets, and manual or semi-automated monitoring to regulate cloud expenditures. Due to their simplicity and low implementation complexity, static techniques are widely adopted in early-stage cloud deployments and environments with relatively stable and predictable workloads.

METHODOLOGY:

Budgeting and Cost Allocation: Where spending limits are predefined for departments, projects, or services, and cloud usage is monitored against these limits. Such budgeting mechanisms help organizations gain financial visibility and prevent uncontrolled spending.

Tagging and Resource Grouping: It support accountability by associating metadata with cloud resources, allowing costs to be traced back to specific teams or applications. These practices are essential for governance and reporting, particularly in enterprise-scale deployments [11].

Scheduling and Instance Shutdown: Configuring resources to automatically shut down during off-peak hours [16].

Cost Alerts and Reporting: Using built-in cloud dash boards such as AWS Cost Explorer, Azure Cost Management, and GCP Billing Reports, which notify users when predefined budget thresholds are exceeded and provide periodic expenditure summaries [6].

Critical Evaluation:

Although static approaches improve cost visibility and financial discipline, they lack adaptability in dynamic cloud environments. Fixed budgets and rule-based alerts do not respond effectively to sudden workload variations, leading to delayed corrective actions and potential cost overruns. These approaches are particularly ineffective for micro services-based, event-driven, or highly elastic applications, where resource demand changes rapidly. Consequently, static techniques are best suited as a foundational governance layer rather than a standalone solution and must be complemented by dynamic or intelligent optimization mechanisms for effective cloud cost management.

Dynamic Approaches

Dynamic approaches to cloud cost management focus on adjusting resource allocation in response to real-time workload variations. Unlike static techniques, these methods continuously monitor system performance and automatically scale resources to balance cost efficiency and application performance. Dynamic approaches are particularly effective in cloud-native environments where demand fluctuates due to varying user activity or event-driven workloads.

Common techniques include:

Auto-Scaling: It enables cloud resources to scale up or down based on predefined performance metrics such as

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

CPU utilization, memory usage, or incoming request rate. Auto-scaling may be implemented either horizontally by adding or removing compute instances or vertically by resizing instance capacities. Studies on cloud-native applications show that well-designed auto-scaling policies can significantly minimize resource underutilization and reduce operational costs while maintaining acceptable performance levels [1].

Spot and Preemptible Instances: A dynamic strategy which exploit unused cloud capacity at significantly reduced prices. Although these instances offer substantial cost savings, their availability is not guaranteed, as they may be terminated without notice. Consequently, this approach is most suitable for fault-tolerant, batch, or non-critical workloads.

Serverless Computing: Using pay-per-execution services like AWS Lambda or Google Cloud Functions to eliminate idle infrastructure costs.

Load-Balancing and Scheduling: Dynamic approaches are particularly beneficial in multi-cloud environments, where load balancing and runtime scheduling can distribute workloads across providers based on cost and performance considerations. Research on runtime microservice re-orchestration demonstrates that dynamically migrating workloads can achieve meaningful cost reductions in multi-cloud systems [3]. Similarly, adaptive orchestration mechanisms enable applications to optimize performance-cost trade-offs across heterogeneous cloud infrastructures [5].

Critical Evaluation:

Despite their advantages, dynamic approaches introduce notable operational challenges. Frequent scaling operations may cause performance instability, increased application latency, and management overhead. Additionally, reliance on spot or preemptible instances increases vulnerability to unexpected resource termination, requiring robust fault-tolerance mechanisms. Provider-specific implementations and configuration complexity further hinder the portability of dynamic cost optimization solutions across multi-cloud platforms. As a result, while dynamic approaches outperform static methods in adaptability, they often benefit from intelligent prediction and automation, highlighting the need for AI and ML-based optimization techniques discussed in the following section.

AI and ML-Based Optimization

Recent advancements in cloud cost management increasingly leverage artificial intelligence (AI) and machine learning (ML) techniques to enable proactive and intelligent cost optimization. Unlike static and dynamic approaches that rely on predefined rules or reactive scaling, AI/ML-based approaches learn from historical usage patterns and adapt resource provisioning strategies to optimize cost–performance trade-offs in evolving cloud environments.

Examples Include

Predictive Cost Modeling: where time-series and regression-based models are employed to forecast future resource demands and associated costs. Statistical approaches such as ARIMA-based forecasting have been applied to estimate short-term cloud usage trends [14], while deep learning models, including Long Short-Term Memory (LSTM) networks, demonstrate improved prediction accuracy for complex and nonlinear workload patterns [15]. These predictive techniques support proactive budgeting, capacity planning, and early detection of potential cost overruns.

Reinforcement Learning for Resource Allocation: RL-based systems continuously interact with the cloud environment and learn optimal provisioning policies by balancing reward functions that incorporate both performance metrics and cost objectives. Graph-based and optimization-driven learning frameworks further enhance cost modeling by capturing interdependencies between cloud resources and services [2].

Intelligent Orchestration: AI-driven approaches are particularly effective in multi-cloud and hybrid environments, where intelligent orchestration mechanisms dynamically distribute workloads across providers

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

based on cost efficiency and performance considerations. Advanced orchestration frameworks enable runtime decision-making to migrate or reallocate micro-services in response to changing cost conditions [3].

FinOps Automation Tools: Platforms like ABACUS [4]: A FinOps Service for Cloud Cost Optimization employ automation and AI to monitor, analyze, and enforce financial governance policies.

Critical Evaluation:

Despite their high adaptability and potential cost savings, AI and ML-based cost optimization techniques face several practical challenges. These systems are heavily dependent on the quality and availability of historical data, making them vulnerable to model bias and performance degradation due to workload drift. Additionally, the limited interpretability of complex learning models reduces transparency in financial decision-making, which can hinder trust and adoption among stakeholders. Integration complexity, increased computational overhead, and data privacy concerns further restrict the large-scale deployment of AI-driven solutions. As a result, AI/ML-based approaches are most effective when combined with robust governance frameworks and carefully designed validation mechanisms.

Table 1: Comparison of Cloud Cost Management Approaches

Approach	Technique	Adaptability	Cost Efficiency	Complexity	Key Limitations
Static	Budgeting, Tagging, Scheduling	Low	Moderate	Low	Poor response to dynamic workloads, delayed cost correction
Dynamic	Auto-scaling, Spot Instances, Serverless	High	High	high	Performance instability, configuration complexity
AI/ML-based	Predictive Modelling, Reinforcement Learning	Very High	Very high	Very high	High data dependency, limited explain ability, integration overhead

Challenges and Limitations

Despite continuous advancements in cloud cost management techniques, organizations still face significant challenges in maintaining effective financial control over cloud resources. These challenges arise from the inherent complexity of cloud environments, diverse pricing models, and the limitations of existing static, dynamic, and AI-driven approaches.

Limited Cost Visibility in Multi-Cloud Environments:

One of the most critical challenges in cloud cost management is achieving unified cost visibility across multi-cloud deployments. Static and dynamic approaches often rely on provider-specific billing systems, which differ in cost metrics and reporting formats. This fragmentation limits centralized governance and hinders real-time identification of cost inefficiencies across platforms [6].

Complex and Unpredictable Pricing Models:

Cloud service providers continuously update pricing structures, offering a wide range of instance types and usage-based billing options. Static budgeting methods are particularly ineffective under such pricing volatility, leading to inaccurate forecasting and budget overruns. Unexpected data egress fees further complicate financial

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

planning and frequently remain hidden until billing cycles are completed [9].

Inefficient Resource Utilization:

Over-provisioning remains a major contributor to inflated cloud costs. While dynamic approaches reduce idle resources through scaling mechanisms, they may cause performance instability when not properly tuned. Manual configuration alone is insufficient for optimizing cost-performance trade-offs in highly dynamic and microservices-based workloads [3].

Lack of Standardization and Interoperability:

The absence of standardized billing APIs and cost management frameworks limits interoperability between cloud providers. This constraint restricts the effectiveness of centralized cost optimization tools and increases reliance on third-party monitoring solutions, which may introduce additional financial and operational overhead [6].

Data Transfer and Hidden Costs

One of the most overlooked aspects of cloud billing [7] is data movement between regions or providers. Transferring data across cloud platforms incurs egress fees, which can represent a significantportionoftotalcloudexpenditure.Thesecostsarenotalwaystransparentinbilling reports, leading to misinformed financial decisions.

Limited Adoption of FinOpsCulture

While FinOps practices aim to bridge the gap between finance and engineering teams, many organizations struggle to implement them effectively. The absence of proper financial governance, accountability frameworks, and skilled personnel results in fragmented cost management efforts. As observed in “ABACUS: A FinOps Service for Cloud Cost Optimization” [4], automation alone cannot replace the organizational mindset and collaboration required for sustainable cost governance.

Integration, Security, and Data Privacy Concerns:

AI and ML-based cost optimization techniques depend heavily on large volumes of operational and financial data. Integrating these systems introduces security and compliance challenges, especially in regulated environments. Inaccurate predictive models or mis-configured automation workflows may unintentionally increase operational costs rather than reduce them [7].

Future Trends and Research Directions

The evolution of cloud cost management is shifting from reactive cost monitoring toward proactive, intelligent, and sustainable optimization [13]. As cloud adoption deepens and workloads diversify, emerging technologies and financial practices are expected to redefine how organizations manage cloud expenditures.

AI-Driven Cost Forecasting and Automation

Artificial intelligence and machine learning are expected to play a central role in next-generation cloud cost management systems. Advanced forecasting models based on time-series analysis and deep learning can enable more accurate prediction of resource demand and associated costs [14][15]. Future research may explore reinforcement learning-based self-optimizing systems capable of autonomously adjusting provisioning strategies in real time while balancing cost, performance, and availability constraints [2].

FinOps Maturity and Cultural Integration

TheFinOps (Financial Operations) movement is gaining traction as a collaborative framework connecting finance, engineering, and operations teams. Next-generation FinOps platforms are expected to evolve beyond simple cost dash boards into autonomous decision systems that enforce budgets, predict cost overruns, and recommend corrective actions dynamically. Future studies could focus on establishing FinOps maturity models that quantify how effectively an organization manages cloud costs, as well as standardized KPIs to measure cost efficiency across industries.

Sustainable and Green Cloud Cost Management

Sustainability is emerging as a dual priority alongside cost efficiency. Cloud providers are beginning to publish carbon footprint metrics for their services, enabling users to make cost energy-aware deployment choices. Research opportunities exist in eco-aware resource scheduling and carbon-efficient workload distribution, which aim to minimize both financial cost and environmental impact. Integrating sustainability metrics into cost-optimization algorithms can drive the development of greener cloud computing frameworks.

Cross-Cloud Interoperability and Unified Billing Standards

As multi-cloud adoption [10] increases, interoperability between providers becomes crucial. Future work should aim at establishing standardized billing APIs and cross-provider data- exchange protocols to provide unified cost visibility. Open-source initiatives and industry consortiums can play a key role in creating these standards, reducing dependency on vendor- specific tools.

Security-Aware Cost Optimization

Cost optimization must evolve alongside security and compliance requirements. Future researchmayexploremodelsthatjointlyoptimizecost,performance,andsecurityposture.For instance, integrating policy-driven cost governance with cloud security management could prevent cost leakage caused by redundant or mis-configured resources.

Integration with Edge and Serverless Architectures

With the rise of edge computing and serverless frameworks, cost management needs to expand beyond centralized cloud environments. Research directions include developing adaptive cost-control mechanisms for distributed, latency-sensitive edge nodes and optimizing serverless billing models based on real-timeevent loads.

CONCLUSION

Cloud Cost Management has emerged as a critical discipline in ensuring the financial sustainability of cloud-based infrastructures. As organizations continue to migrate toward multi-cloud and hybrid environments, managing and optimizing cloud expenditure has become increasingly complex. This review paper has analyzed the evolution of cost management techniques, ranging from static budgeting approaches to dynamic and AI-driven optimization strategies.

While existing tools and frameworks offer partial solutions, major challenges such as cost visibility, unpredictable pricing models, and lack of standardization still persist. The findings from recent studies indicate that automation, intelligence, and financial collaboration are essential to overcoming these limitations. The integration of FinOps principles with AI-based cost prediction and orchestration represents a promising pathway toward intelligent, autonomous cost control systems.

Looking ahead, the future of Cloud Cost Management lies in achieving a balance between economic efficiency, operational performance, and environmental sustainability. Research efforts should continue to focus on developing interoperable, adaptive, and energy-efficient cost optimization [12] frameworks that align technological innovation with organizational goals.

Ultimately, effective cloud cost management is not only a matter of reducing expenses but also a strategic enabler of innovation, agility, and long-term digital resilience.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XI, November 2025

REFERENCE

1. N. Ramesh,A.Banerjee,andV.K.Singh,"Auto-scalingApproachesforCloud-Native Applications: ASurvey and Taxonomy," *Sensors*, vol. 24, no. 5, pp. 1–18, Mar. 2024.
2. T.M. Nguyenand Y. Lee, "Cost Modellingand Optimisation for Cloud:AGraph-Based Approach,", Journal of Cloud Computing, vol. 13, no. 2, pp. 115 132, 2024
3. P. GhoshandL. Chen,"CostMinimizationinMulti-CloudSystemswithRuntime Microservice Re-orchestration," arXiv preprint, arXiv:2403.09121, 2024. [Online].
4. K.Sharma,D.Raj,andJ.K.Verma,"ABACUS:A FinOps Service for Cloud Cost Optimization,",arXiv preprint, arXiv:2501.06732, 2025.
5. M.Zhaoand, R.Gupta,"AdaptiveOrchestrationforPerformance–CostOptimizationin Multi-Cloud,", SSRN Electronic Journal, 2020.
6. K. Bedi et al., "Unified Cost Visibility in Multi-Cloud," IEEETrans. Cloud Comput., vol. 11, pp. 89–101, 2023.
7. Singh et al., "Blockchain for Cloud Billing Audit," IEEEAccess, vol. 12, pp. 7823–7836, 2024.
8. M. Armbrust et al., "A View of Cloud Computing," Commun.ACM, vol. 53, no. 4, pp. 50–58, 2010.
9. E. Walker, "The Real Cost of a CPU Hour," Computer, vol. 42, no.4, pp. 35–41, 2011.
10. Khajeh-Hosseini et al., "The Cloud Adoption Toolkit," Softw.Pract. Exp., vol. 42, no. 4, pp. 447–465, 2012.
11. R. Buyya et al., "Market-Oriented Cloud Resource Allocation,"Future Gener. Comput. Syst., vol. 26, pp. 1012–1023, 2013.
12. Beloglazov and R. Buyya, "Energy-Efficient Resource Allocation," Future Gener. Comput. Syst., vol. 28, pp. 755–768, 2012.
13. S. Jain and R. Gupta, "Policy-Based Cloud Optimization," IEEE Cloud Comput., vol. 6, pp. 40–48, 2019.
14. X. Li and Y. Chen, "ARIMA Forecasting in Cloud Environments,"Future Gener. Comput. Syst., vol. 107, pp. 509–519, 2020.
15. J. Zhao et al., "LSTM-Based Cost Forecasting," IEEE Access, vol. 9, pp. 18123–18135, 2021.
16. M. Yadav, A. Mishra, Energy-efficient workflow scheduling using dynamic task clustering for sustainable cloud computing. Discov Computing, 28, article id 201, 2025. https://doi.org/10.1007/s10791-025-09712-0.