

GLOBAL JOURNAL OF ADVANCED ENGINEERING TECHNOLOGIES AND SCIENCES**CLOUD COST EFFICIENCY STRATEGIES AND IMPLEMENTATION WITH A FOCUS ON AWS****Akshay Jha*¹, Arpit Solanki*²**^{*1}Student, Dr. A.P.J. Abdul Kalam University, Indore, M.P., India^{*2}Assistant Professor, Dr. A.P.J. Abdul Kalam University, Indore, M.P., India

ABSTRACT

Scalability, agility and global accessibility are the drivers of modern digital transformation which make use of cloud computing to provide benefits on industries. Nevertheless, the increased adoption of cloud has additionally created a huge challenge in terms of cost especially when it comes to enterprises using extensive multi-cloud and AWS infrastructure. In this paper, the cost-efficiency strategies in clouds are explored, the services that are considered the best in the industry are analyzed, and the best-practice approaches to implementing the cost-reduction strategies are identified. This work is an in-depth, evidence-based examination of the cost optimization based on analyzing current literature, cloud migration models and templates, resource allocation systems, and server less architecture. Results indicate that cloud governance, rightsizing, reserved instances, server less compute, AI-based optimization, workload visibility, and best practices could help enterprises to save 25-60 percent in cloud costs. This paper is relevant to the practical knowledge and proven practices of organizations with the aim of enhancing cloud effectiveness, as well as through performance and ROI maximization.

KEYWORDS: Cloud Computing, AWS

INTRODUCTION

Cloud computing has become a cornerstone of modern digital transformation by enabling scalable, on-demand access to computing resources with minimal upfront investment. Over the past decade, it has evolved from a supporting IT solution into a dominant computing paradigm that powers enterprise applications, data analytics, artificial intelligence, and Internet of Things (IoT) systems. Leading cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have largely replaced traditional on-premises infrastructures by offering flexible and usage-based pricing models.

Despite these advantages, organizations increasingly face challenges related to uncontrolled cloud expenditure. Research indicates that a substantial portion of cloud spending is wasted due to over-provisioning, underutilized resources, inefficient scaling mechanisms, and the lack of effective governance and monitoring. The complexity of cloud pricing models, combined with decentralized resource provisioning, further exacerbates cost inefficiencies, particularly in large and multi-team environments. AWS, as the leading cloud service provider, offers a wide range of instance types, pricing options, and automation features designed to optimize costs. However, selecting the most cost-efficient configuration remains difficult without continuous monitoring and data-driven decision-making. Consequently, cloud cost optimization has emerged as a critical concern for organizations seeking to maximize return on investment while maintaining performance and reliability.

This paper investigates cloud cost efficiency with a focus on AWS by comparing historical pricing data with real-time Spot Instance experimentation. The study demonstrates how newer instance generations and appropriate pricing models can significantly reduce cloud computing costs and provides insights into practical strategies for sustainable cloud cost optimization.

RELATED STUDY

This section outlines, some noteworthy contributions has been reported

Several research works have addressed cloud cost optimization, resource utilization, and pricing efficiency in cloud computing environments. Ekwe-Ekwe and Barker (2018) conducted a detailed analysis of Amazon EC2 Spot Instance pricing across multiple geographical regions and demonstrated that region selection plays a critical role in achieving cost stability and savings, with reductions of up to 90% compared to On-Demand pricing for fault-tolerant workloads.

Jonnalagadda (2024) investigated cost-effective cloud resource utilization and identified over-provisioning and idle resources as primary contributors to cloud cost inefficiency. The study emphasized rightsizing, predictive

auto-scaling, serverless architectures, and AI-driven automation as key techniques for reducing unnecessary cloud expenditure. Similarly, Mahida (2022) reviewed dynamic resource allocation and machine learning-based provisioning approaches to balance performance and cost in cloud environments.

Energy-efficient and sustainable cloud cost optimization has been explored by Giacobbe et al. (2023) and Chiang, Ouyang, and Hsu (2014). Their studies proposed green computing strategies such as energy-aware scheduling, VM consolidation, and adaptive resource control to minimize energy consumption and indirectly reduce cloud service costs.

Tang (2021) presented a reliability-aware and cost-efficient workflow scheduling strategy for multi-cloud systems, showing that intelligent workload distribution can lower execution costs without sacrificing reliability. Ahmed et al. (2021) analyzed serverless computing models and highlighted AWS Lambda as a cost-efficient solution for event-driven and intermittent workloads by eliminating idle compute costs.

Although these studies provide strong theoretical and practical foundations, limited research integrates historical cloud pricing data with real-time billing experiments. This study addresses this gap by combining historical AWS Spot pricing with live experimental cost analysis to evaluate practical cloud cost efficiency.

METHODOLOGY

This study adopts a comparative experimental methodology to evaluate cloud cost efficiency by analyzing historical AWS Spot Instance pricing data alongside real-time experimental billing observations. The objective is to assess whether newer-generation Amazon EC2 instances provide measurable cost advantages over legacy instance families when deployed using Spot pricing. Historical pricing data for the legacy m3.large instance were obtained from previously published research, which reported an average Spot price of USD 0.088 per hour for the EU region. Since the m3.large instance family has been deprecated and is no longer available for deployment, this dataset serves as a baseline reference for comparison. For real-time experimentation, a modern m7i-flex.large EC2 Spot Instance—configured with the same computational capacity (2 vCPUs and 8 GiB RAM)—was deployed in the us-east-1 region. A controlled CPU-intensive workload was executed continuously for one hour to ensure consistent resource utilization. Runtime behavior and execution continuity were monitored using Amazon CloudWatch to verify uninterrupted operation. After completion, the exact billed cost was extracted from the AWS Billing Dashboard and cross-validated using AWS Cost Explorer to ensure pricing accuracy. The collected real-time cost data were compared with historical pricing using a standardized cost improvement formula to quantify percentage savings. This methodology provides a reproducible and data-driven framework for evaluating cloud cost efficiency and identifying economically optimal EC2 instance configurations.

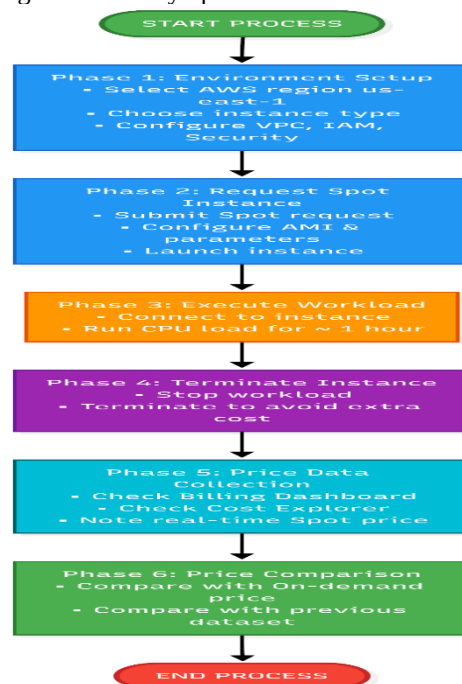


Fig: 1 Experimental Workflow for Instance Cost Comparison

RESULT ANALYSIS

The real-time Spot experiment on the *m7i-flex.large* EC2 instance recorded an hourly cost of **USD 0.03**, compared to the historical *m3.large* Spot price of **USD 0.088**. With identical configurations (2 vCPUs, 8 GiB RAM), the results indicate a **65.9% cost reduction**. This confirms that newer EC2 instance generations provide significantly better cost efficiency under Spot pricing.

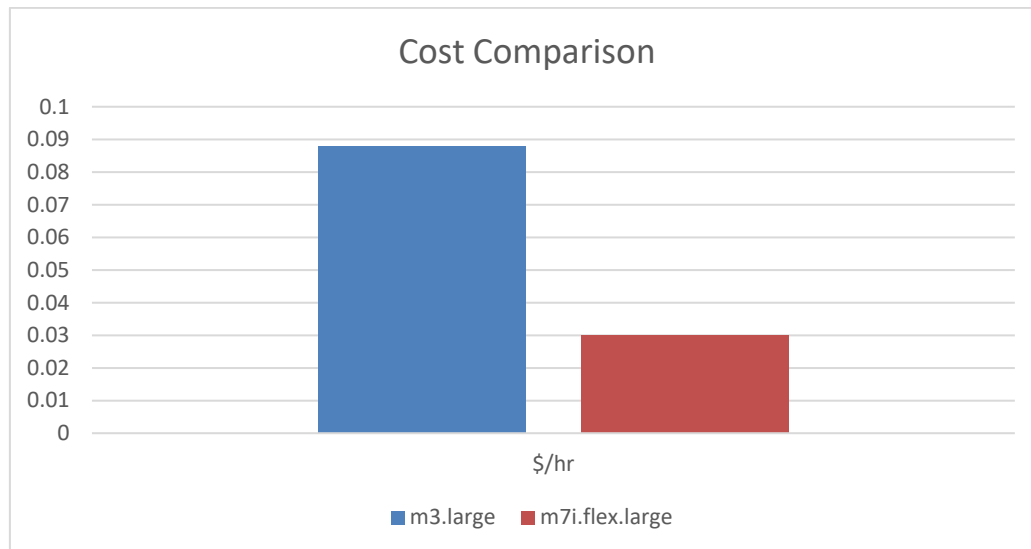


Fig: 2 Cost Comparison between Historical and Experimental Pricing

CONCLUSION AND FUTURE SCOPE

This study analyzed cloud cost efficiency in Amazon Web Services by comparing historical EC2 Spot pricing with real-time experimental billing data. The results demonstrate that newer-generation EC2 instances, when deployed using Spot pricing, provide substantial cost advantages over legacy instance families while maintaining equivalent performance. A cost reduction of approximately 65.9% was observed, validating the effectiveness of the proposed model. The findings emphasize that continuous evaluation of instance generations and real-time pricing is essential for achieving sustainable cloud cost optimization.

Future research can extend this work by evaluating multiple EC2 instance families across different AWS regions and longer execution durations to capture broader pricing patterns. Incorporating automated instance selection, predictive Spot price forecasting, and AI-based optimization techniques may further improve cost efficiency. Additionally, integrating the proposed model with FinOps frameworks and multi-cloud environments can enhance its applicability for large-scale and enterprise cloud deployments.

REFERENCES

- [1] N. Ekwe-Ekwe and A. Barker, "Location, Location, Location: Exploring Amazon EC2 Spot Instance Pricing Across Geographical Regions," *Proc. 18th IEEE/ACM Int. Symp. Cluster, Cloud and Grid Computing (CCGrid)*, Washington, DC, USA, 2018, pp. 370–373.
- [2] A. M. C. Jonnalagadda, "Cost-Effective Cloud Solutions: Optimizing Resource Utilization and Expenditure," *Int. J. Res. Comput. Appl. Inf. Technol. (IJRCAIT)*, vol. 7, no. 2, pp. 1–10, 2024.
- [3] S. Deochake, "Cloud Cost Optimization: A Comprehensive Review of Strategies and Case Studies," *arXiv preprint arXiv:2307.12479*, 2023.
- [4] X. Tang, "Reliability-Aware Cost-Efficient Scientific Workflow Scheduling Strategy on Multi-Cloud Systems," *IEEE Trans. Cloud Comput.*, vol. 10, no. 4, pp. 2909–2919, 2021.
- [5] D. Yuniarto, "Implementing Cloud Computing in Companies to Increase Business Efficiency," *Jurnal Info Sains: Informatika dan Sains*, vol. 13, no. 2, pp. 633–639, 2023.
- [6] S. D. Pasham, "AI-Driven Cloud Cost Optimization for Small and Medium Enterprises," *The Computertech*, pp. 1–24, 2017.
- [7] H. Sharma, "Effectiveness of Cloud Security Posture Management in Multi-Cloud Environments," *Int. J. Comput. Sci. Eng. Res. Dev. (IJCSERD)*, vol. 10, no. 1, pp. 1–18, 2020.



- [8] Y. J. Chiang, Y. C. Ouyang, and C. H. Hsu, "An Efficient Green Control Algorithm in Cloud Computing for Cost Optimization," *IEEE Trans. Cloud Comput.*, vol. 3, no. 2, pp. 145–155, 2014.
- [9] A. Mahida, "Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency," *J. Artif. Intell. Cloud Comput.*, vol. 1, no. 232, pp. 2–4, 2022.
- [10] V. H. Pardeshi, "Cloud Computing for Higher Education Institutes: Architecture, Strategy and Recommendations for Effective Adaptation," *Procedia Econ. Finance*, vol. 11, pp. 589–599, 2014.
- [11] N. Ahmed, M. E. Hossain, S. S. I. Rishad, N. N. Rimi, and M. I. Sarkar, "Serverless Architecture: Optimizing Application Scalability and Cost Efficiency in Cloud Computing," *BULLET: J. Multidisciplinary Sci.*, vol. 1, no. 6, pp. 1366–1380, 2021.
- [12] I. Naseer, "AWS Cloud Computing Solutions: Optimizing Implementation for Businesses," *Stat., Comput. Interdiscip. Res.*, vol. 5, no. 2, pp. 121–132, 2023.
- [13] Amazon Web Services, "Amazon EC2 Instance Types," AWS Documentation, 2024. [Online]. Available: <https://aws.amazon.com/ec2/instance-types/>
- [14] Amazon Web Services, "Amazon EC2 Pricing," AWS Documentation, 2024. [Online]. Available: <https://aws.amazon.com/ec2/pricing/>