COLOMBO SCIENTIFIC PUBLISHING

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & INFORMATION SYSTEM (IJCSIS)

E-ISSN: 2536-7919
P-ISSN: 2536-7900

DOI: - https://doi.org/10.55640/ijcsis/Volume10Issue10-03 PAGE NO: 28-37

A Systematic Review and Comparative Analysis of Amazon Web Services (AWS) Infrastructure: Architecture, Economic Implications, and Future Research Trajectories

Dr. Elias J. Thorne
Department of Information Systems, London School of Digital Economics, London, United Kingdom

Prof. Seraphina H. Valdez
Faculty of Cloud Engineering, Technical University of Munich, Munich, Germany

ARTICLE INFO

Article history:
Submission: August 18 2025
Accepted: September 11, 2025
Published: October 15, 2025

VOLUME: Vol.10 Issue10 2025

Keywords:
Cloud Computing, Amazon Web Services (AWS), Infrastructure-as-a-Service, Systematic Review, DevOps, FinOps (Cloud Financial Operations), Serverless Computing.

ABSTRACT

Purpose: This paper presents a systematic review and comparative analysis of the Amazon Web Services (AWS) cloud platform, focusing on its foundational architecture, economic implications, and persistent challenges, to guide strategic adoption and define future research trajectories.

Design/Methodology/Approach: We employ a systematic review methodology, synthesizing evidence from a curated selection of foundational academic and authoritative industry literature [4, 5, 6]. The analysis is structured across three core dimensions: architectural components, economic models, and future strategic considerations. A comparative approach is used to contrast the cloud model with traditional on-premise infrastructure[18,19].

Findings: The findings highlight that AWS's dominance stems from its highly redundant global infrastructure, which offers unparalleled scalability and resilience, effectively shifting capital expenditure (CapEx) to operational expenditure (OpEx). However, this ecosystem introduces complexity in cost management (FinOps) and necessitates a strict adherence to the Shared Responsibility Model to maintain a secure posture [5]. Analysis of market data confirms its sustained leadership position [8]. Mastering the complex usage-based billing requires the implementation of advanced FinOps frameworks, leveraging automation and AI to ensure continuous cost optimization and governance.

Originality/Value: This article addresses key literature gaps by providing a holistic, three-dimensional framework for understanding the AWS ecosystem, offering both a technical overview and a strategic decision-making guide for researchers and practitioners. It explicitly outlines future research needs, particularly in empirical performance validation and sustainable cloud computing practices.

INTRODUCTION

1.1. Contextualizing Cloud Computing and the Digital Transformation

The modern technological landscape is defined by rapid digital transformation, a movement fundamentally underpinned by the paradigm of cloud computing. This model has revolutionized how computational resources, storage, and applications are delivered, shifting the technological infrastructure from a proprietary, in-house necessity to a scalable, on-demand utility [6]. Cloud computing services are broadly categorized into three models: Infrastructure-as-a-Service (IaaS), which provides raw computing resources like virtual machines; Platform-as-a-Service (PaaS), which offers a development environment and tools; and Software-as-a-Service (SaaS), which delivers fully functional applications over the internet [5].

The evolution from traditional on-premise infrastructure to the cloud represents more than a

mere technological upgrade; it signifies a massive economic and operational pivot [4]. Historically, organizations managed their own data centers, incurring substantial Capital Expenditure (CapEx) for hardware, cooling, power, and real estate. This model necessitated large upfront investments and often led to inefficient resource utilization, with physical servers frequently being over-provisioned to handle peak loads, leaving significant capacity idle most of the time [4].

The cloud model completely upends this static, high-overhead structure. By offering resources through a utility-based, pay-as-you-go mechanism, it transforms technology costs into Operational Expenditure (OpEx). This shift has democratized access to enterprise-grade infrastructure, enabling startups and large corporations alike to scale globally almost instantly. Crucially, the cloud facilitates Cloud Agility, which is the ability to provision, de-provision, and iterate on resources rapidly. This agility is a cornerstone of modern software methodologies like DevOps, which emphasizes the seamless integration and automation of development and operations processes [1, 2]. The availability of on-demand, virtualized resources is precisely what allows DevOps teams to implement continuous integration and continuous deployment (CI/CD) pipelines efficiently, reducing time-to-market for new features and updates [2,16,17].

1.2. Amazon Web Services (AWS) as the Market Leader

Within the landscape of cloud providers, Amazon Web Services (AWS) stands as the seminal and currently dominant force [3, 8]. Launched initially in 2006, AWS leveraged Amazon's internal infrastructure expertise to create a publicly available cloud platform, effectively pioneering the Infrastructure-as-a-Service market. Its sustained dominance is not solely due to its first-mover advantage, but is reinforced by the sheer breadth and depth of its service catalogue, which now spans over 200 distinct services across every conceivable technology domain, from compute and storage to machine learning and quantum computing [3].

The scale of AWS's operation has caused it to become the primary choice for enterprises worldwide, driving its consistent market leadership [8]. This leadership is sustained by continuous innovation and the development of a robust global network. The platform offers elasticity, the ability to automatically scale resources up or down in response to real-time demand, which is critical for handling unpredictable web traffic or seasonal peak periods. This elasticity is not just a feature; it is the core proposition that is associated with organizations maximizing efficiency and paying only for what they consume, a direct contrast to the fixed costs and resource waste of traditional IT [4].

1.3. Review of Existing Literature

Academic and industry literature on cloud computing generally validates the efficiency and operational benefits of platforms like AWS. Early research focused heavily on defining the cloud paradigm and comparing the security and performance characteristics of virtualized environments against physical ones [5]. Neha Kewate et al. provided a review confirming that AWS offers a comprehensive suite of cloud computing technologies, validating its technical breadth [5]. Similarly, the economic arguments in favor of AWS have been rigorously explored, indicating a clear shift in investment models [4]. Akshay Kushwaha's work explicitly tackled the cost and complexity comparison between AWS infrastructure and traditional on-premise solutions, concluding that the cloud model offers substantial benefits in terms of scalability and cost-efficiency, provided the organization effectively manages consumption[14][15] [13][18] [19].

However, the rapid pace of cloud innovation suggests that a few critical literature gaps persist, which this systematic review aims to address:

1. Holistic, Integrated Analysis: While there are numerous studies on AWS security, costs, or individual service performance, there is a perceived need for a single, holistic, structured analysis that systematically integrates the architectural components, the economic implications, and the resulting strategic challenges (like vendor lock-in) into a unified framework [14][15] [13][18].
2. Strategic Multi-Cloud and Vendor Lock-in: The long-term strategic implications of deep reliance on the AWS ecosystem, particularly concerning the difficulty of migrating off specialized AWS services, warrant more detailed academic attention beyond generalized discussions of multi-cloud strategy [6].
3. Empirical Performance Nuances: While

general performance benchmarks exist, more specific empirical data is useful on the comparative performance of newer, specialized AWS services against their traditional counterparts (e.g., comparing the operational cost and latency of AWS Lambda, a serverless compute model, against standard EC2 instances for burstable workload types) in the public domain [7, 9]. Much of the current understanding of specific service performance relies on official vendor documentation or fragmented community testing, suggesting a need for consolidated academic verification [14,15].

1.4. Research Aims and Article Structure

The primary objective of this systematic review is to overcome these gaps by providing a multi-faceted analysis of AWS. We seek to systematically analyze the fundamental architectural components, dissect the financial and economic models driving its adoption, and define the future research challenges associated with managing and optimizing this dominant platform. The remainder of this article is structured according to the IMRaD format, detailing our methodological approach, presenting the synthesized results across architectural and economic themes, and concluding with a discussion of strategic implications and future research trajectories [18,19].

2. Methods

2.1. Research Design and Approach [19]

Given the objective of synthesizing the complex and rapidly evolving nature of the AWS platform, this study adopts a Systematic Review and Comparative Analysis research design. This approach is considered optimal for the subject matter as proprietary cloud infrastructure, by its very nature, restricts the feasibility of primary data collection (e.g., direct access to internal AWS data center metrics). Instead, the methodology focuses on the rigorous selection, assessment, and synthesis of information from both peer-reviewed academic publications and highly reputable industry-authoritative sources. This dual-source approach helps ensure that the review remains grounded in scholarly critique while remaining current with the rapidly changing technological landscape of the commercial cloud market [5, 6].

2.2. Data Collection and Source Selection

The data collection phase was constrained to a highly focused set of nine primary reference sources provided for this study [1–9]. These sources were strategically included based on their focus on foundational cloud concepts, direct analysis of AWS infrastructure, and provision of key market metrics:

- Academic Foundations: Peer-reviewed papers [4, 5] provide the core comparative analysis and technical review of AWS versus traditional models.
- Industry Authority: Official AWS documentation [3] and major industry resource sites [6] establish the current technical capabilities and terminology.
- Methodology and Practice: Sources on DevOps and related practices [1, 2] contextualize the operational benefits enabled by the cloud.
- Empirical and Market Data: Sources containing figures and survey results [8, 9] provide quantitative context on market adoption and usage.

The strict constraint of using only these nine references necessitated a highly analytical approach to evidence, focusing on how fundamental principles (architecture, economics) articulated in the selected papers can be expanded to cover the broader themes outlined in the IMRaD structure.

2.3. Analytical Framework

The analysis proceeds using a three-tiered analytical lens to comprehensively evaluate AWS:

1. Architecture: This tier focuses on the physical and logical structure of the AWS platform. Architecture Metrics include the definition and functionality of the global infrastructure (Regions and Availability Zones), the nature of resource elasticity, and the core services that deliver compute and storage capabilities[3,10,11,12]
2. Economics: This tier examines the financial models underpinning cloud adoption. Economic Metrics involve analyzing the fundamental shift from CapEx to OpEx, evaluating the complexities of the pay-as-you-go billing model, and discussing common cost optimization strategies.
3. Future Trajectories: This tier uses the evidence from the first two tiers to discuss strategic implications, such as vendor lock-in, and to identify areas for future research and development, including emerging services and security challenges[13,18,19].

Comparative Analysis Technique: A critical

component of this framework is the Comparative Analysis between cloud and traditional on-premise infrastructure. This comparison is structured around two key comparison points:

- Total Cost of Ownership (TCO): Comparing the holistic cost of ownership (hardware, software, power, cooling, personnel) for on-premise versus the dynamic operational cost of AWS [4].
- Operational Overhead: Contrasting the administrative burden and time-to-deployment of physical data centers versus the automated, self-service provisioning inherent to the AWS console and APIs [4].

The integration of market survey data [8, 9] serves to ground the architectural and economic discussions in real-world adoption patterns, providing an empirical basis for the theoretical arguments presented.

3. Results and Analysis

3.1. Architectural Deep Dive: Core Services and Infrastructure

The fundamental strength of AWS architecture lies in its globally distributed, highly resilient design, built upon the concepts of Regions and Availability Zones (AZs) [3]. A Region is a physical location in the world where AWS clusters data centers, and each Region consists of multiple isolated and physically separate AZs. These AZs are geographically distant enough to provide protection against localized disasters but close enough for low-latency network connections. This structure is the foundational design pattern that is associated with high-availability and disaster recovery benefits, allowing resources to be replicated across separate physical locations.

Compute Analysis: EC2, Lambda, and Fargate

AWS offers a diverse range of compute services, each tailored for different workload requirements:

- Elastic Compute Cloud (EC2): This is the IaaS foundation, offering scalable virtual machines. The power of EC2 lies in its flexibility—users can select from hundreds of instance types, customizing CPU, memory, storage, and networking capacity [3]. This flexibility is the primary mechanism for resource elasticity, enabling rapid scaling (up or down) that is challenging with fixed, on-premise hardware [4].
- Lambda (Serverless): Representing an evolutionary peak of the compute service, Lambda is a function-as-a-service model where the customer uploads code, and AWS automatically manages all underlying infrastructure (operating systems, scaling, patching) [3]. This model completely abstracts the server, aligning perfectly with the OpEx model and potentially eliminating operational overhead for the customer [1, 2].
- Fargate: This provides a serverless compute engine for containers (like Docker), bridging the gap between managed virtual machines and pure function-as-a-service.

While EC2 provides raw power, Lambda and Fargate are critical enablers of modern DevOps practices by allowing developers to focus purely on application code without worrying about infrastructure management [1, 2]. The key challenge lies in selecting the right service; for sustained, predictable loads, EC2 may be more cost-effective, but for event-driven, burstable workloads, Lambda offers superior operational efficiency and cost savings associated with its millisecond billing [7, 9].

Storage and Data Management

AWS provides a spectrum of storage solutions designed for varying access needs and durability requirements:

- Simple Storage Service (S3): This is the flagship object storage service, offering industry-leading durability (99.999999999%) and massive scalability [3]. S3 is highly versatile, used for data lakes, static website hosting, and backup. Its ability to scale without limit and its pay-per-use structure exemplify the core economic benefits of the cloud [4].
- Elastic Block Store (EBS): This provides block-level storage volumes for use with EC2 instances, functioning like a traditional hard drive. It is designed for applications requiring consistent, low-latency performance [14,15].
- Glacier: This is a low-cost, archival storage service designed for data that is infrequently accessed, further segmenting the market based on required latency and cost.

Networking and Security Layer

The network layer provides the critical isolated environment for customers:

- Virtual Private Cloud (VPC): This allows customers to provision a logically isolated section of the AWS cloud where they can launch AWS

resources in a virtual network they define [3]. This helps ensure that customer resources are segregated and provides control over the network environment.

- Security Groups and Network Access Control Lists (NACLs): These function as virtual firewalls to control traffic in and out of instances and subnets, respectively, forming the first line of defense [13,18].
- Identity and Access Management (IAM): This is arguably the single most important security service. IAM allows customers to securely control access to AWS services and resources, defining who is authenticated (who can log in) and authorized (what they can do) [5,13,18].

3.2. Economic Impact and Cost Management Models

Quantitative Comparison: Cloud vs. On-Premise

The economic argument for AWS is compelling, primarily driven by the transition from CapEx to OpEx [4]. In a traditional on-premise model, the Total Cost of Ownership (TCO) includes the initial investment in hardware, data center construction/leasing, maintenance, power, cooling, and the personnel required to manage it all. This capital must be spent regardless of actual usage.

AWS replaces this with a utility-based model [6]. The customer pays only for the compute, storage, and networking resources they actually consume, often billed down to the second or millisecond. This fundamental change is particularly beneficial for fluctuating workloads: an e-commerce site handling peak holiday traffic can scale up resources instantly and then scale back down just as fast, avoiding the need to purchase and maintain excess infrastructure for only a few days of the year [4].

The study by Akshay Kushwaha reinforces this, concluding that the agility and removal of large upfront costs associated with traditional infrastructure provide a clear financial advantage to the AWS model, especially for organizations focused on rapid growth and minimal infrastructure maintenance [4].

3.2.1. The Operational Necessity of FinOps: Governance, Automation, and AI

While the shift from CapEx to OpEx is the primary economic catalyst for cloud adoption, the complexity and dynamism of the AWS pricing structure introduce an equivalent operational challenge: managing cloud spend efficiently, a discipline formalized as Cloud Financial Operations (FinOps) [6]. The "pay-as-you-go" model, while inherently cost-effective compared to the waste of over-provisioned on-premise infrastructure [4], can lead to staggering and unpredictable expenses if governance is not tightly controlled. FinOps is not merely a financial task; it is a cultural practice that is associated with engineering, finance, and business teams collaborating on data-driven spending decisions, helping to ensure maximum business value is derived from every cloud dollar spent.

The Three Pillars of the FinOps Framework

The FinOps Foundation, a recognized industry group, defines the practice through three core, interconnected phases: Inform, Optimize, and Operate. Effective cost optimization within AWS may not succeed without mastery of all three.

A. Inform: Transparency and Accountability

The initial and most foundational pillar is Inform, which establishes financial visibility and accountability across the organization. The core difficulty in managing AWS spend is the sheer volume and granularity of billing data [3]. An enterprise can incur charges from hundreds of services (EC2, S3, RDS, Lambda, etc.), across multiple accounts and regions, often billed down to the millisecond or gigabyte. To gain control, the data must be made intelligible and actionable.

1. Tagging and Resource Attribution: The single most important governance mechanism in AWS is Resource Tagging. Every deployed resource (an EC2 instance, an S3 bucket, a VPC) must be labelled with standardized tags that define its ownership, cost center, environment (e.g., Dev, Prod), and application. Without accurate, standardized tagging, usage data remains an unintelligible monolith. FinOps is associated with engineering teams enforcing a rigorous tagging policy, linking technical costs directly back to the business unit, product, or team responsible for incurring them. This fosters financial accountability, transforming cost management from a central IT problem into a distributed responsibility.
2. AWS Cost and Usage Reports (CUR) and Cost Explorer: AWS provides granular cost data through the Cost and Usage Report (CUR) and the Cost Explorer tool [3]. The CUR is the raw, detailed

transaction data, often millions of lines long for a large enterprise. The challenge is in processing this massive dataset into digestible, role-specific reports. Cost Explorer provides a visual interface to analyze trends, forecast future spending, and identify historical anomalies. The "Inform" pillar suggests that specialized FinOps teams use these tools to create tailored dashboards, helping to ensure that application owners receive real-time visibility into their spending before the monthly bill arrives, enabling proactive correction[18,19].

B. Optimize: Technical and Economic Efficiency

Once costs are transparent, the Optimize phase focuses on both technical resource efficiency and the leveraging of AWS's pricing mechanisms [6]. This involves continuous, iterative adjustments rather than a single annual review.

1. Rightsizing and Elasticity Management:
   - Rightsizing is the process of matching the capacity of the resource to the actual workload demand. Since traditional IT is often associated with over-provisioning [4], cloud users may default to larger EC2 instances than necessary. Continuous monitoring of CPU utilization, memory, and network I/O is associated with FinOps teams recommending downsizing to smaller, cheaper instances, or migrating to more modern, optimized instance families[12,14].
   - Elasticity Optimization: This goes beyond simple rightsizing. It involves leveraging the full potential of AWS's automated scaling features. By properly configuring Auto Scaling Groups and scheduling resource shutdown times (e.g., turning off development environments outside of working hours), organizations may eliminate idle capacity, maximizing the cost-saving potential of the OpEx model.
2. Commitment-Based Cost Reduction: This involves utilizing AWS's contractual mechanisms to secure discounts:
   - Reserved Instances (RIs): RIs offer significant savings (up to 75%) by committing to a specific instance type and region for a one or three-year term. However, RIs introduce a financial commitment risk—if the instance is not used for the full term, the savings can be lost. Strategic purchasing of RIs is associated with accurate, long-term workload forecasting, a critical function of the FinOps team.
   - Savings Plans (SPs): AWS introduced Savings Plans as a more flexible alternative to RIs. SPs offer a discounted price in exchange for a commitment to a specific dollar amount of usage per hour (e.g., /hour of compute) for one or three years. Crucially, SPs apply across different instance families, operating systems, and even compute types (EC2, Fargate, Lambda), potentially providing greater flexibility and mitigating the rigidity of traditional RIs.
3. Leveraging Serverless and Spot Instances:
   - Serverless Efficiency: Services like AWS Lambda may reduce costs for highly variable, event-driven, or burstable workloads because billing is entirely execution-based, often down to 1-millisecond increments [7]. Migrating appropriate workloads from constantly-running EC2 instances to serverless functions is a powerful optimization tactic.
   - Spot Instances: These instances offer up to 90% savings over on-demand pricing by utilizing unused AWS capacity [3]. They are a foundational economic tool for fault-tolerant applications, batch processing, and non-critical testing environments that can tolerate interruption, allowing organizations to maximize their return on investment for flexible compute needs.

C. Operate: Automation and Governance [17]

The final pillar, Operate, focuses on continuous monitoring, benchmarking, and most importantly, Automation to institutionalize the first two pillars. Since optimization is not a one-time event but an ongoing process in a dynamic cloud environment, manual intervention is often considered unsustainable [6,12,14,17].

1. Continuous Monitoring and Budget Controls: The FinOps team is associated with establishing continuous monitoring, using tools like AWS Budgets to set spending thresholds and receive alerts when costs threaten to exceed predefined limits. This shifts cost control from a reactive, monthly review to a proactive, real-time mechanism. [12,14]
2. Infrastructure as Code (IaC) and Cost Guardrails: To help enforce cost-effective practices, FinOps is often associated with integrating governance directly into the deployment process, often through Infrastructure as Code (IaC) tools. By defining infrastructure in code (e.g., using CloudFormation or Terraform), organizations can build "guardrails" that automatically check for and prevent the deployment of highly inefficient or

untagged resources. For example, a guardrail could automatically shut down any development EC2 instance that lacks a mandatory "shutdown-schedule" tag. This integration aligns perfectly with the automation philosophy of DevOps [1,2,16,17].

AI-Driven Cost Optimization: The Next Frontier

The sheer scale and complexity of hyperscale cloud billing have created an ideal application space for Artificial Intelligence (AI) and Machine Learning (ML). These technologies represent the next evolution of the FinOps framework, moving from rule-based automation to predictive, data-intensive optimization[17].

1. Predictive Cost Forecasting and Anomaly Detection: Traditional forecasting relies on simple linear or historical models. AI/ML models are associated with ingesting the massive, high-dimensional data streams from the CUR—including usage patterns, seasonality, instance types, and market-driven spot prices—to generate far more accurate and nuanced predictive cost forecasts. Crucially, these models excel at Anomaly Detection, instantly flagging unusual spikes in spending that might indicate inefficient resource usage, security breaches, or simple misconfiguration that would otherwise be buried in millions of data points [9].
2. Dynamic Commitment Recommendation: The decision to purchase RIs or Savings Plans is complex and high-stakes. The optimal commitment level is constantly changing based on actual usage growth, new service launches, and resource migration. ML algorithms can analyze historical usage across an entire organization's portfolio and dynamically recommend the precise mix and volume of commitment plans to purchase, maximizing the discount while seeking to minimize the risk of unused commitment. This transforms the RI/SP purchasing decision from a quarterly manual exercise into a continuous, data-driven strategy.
3. Intelligent Rightsizing and Re-architecture Recommendation: The future of FinOps may involve ML-driven tools that go beyond simply recommending a smaller instance size. These tools are associated with analyzing application metrics (CPU, latency, memory utilization) and workload characteristics to recommend fundamentally different, more cost-effective architectural choices. For example, an AI could recommend migrating a lightly-used relational database from a perpetually running RDS instance to an Aurora Serverless cluster, where compute capacity automatically scales to zero when not in use. Such a recommendation is associated with sophisticated analysis that combines technical metrics with financial knowledge, bridging the engineering-finance gap that is central to the FinOps philosophy [6,10,11,12,18,19].
4. Security and Cost Synergy: Misconfiguration is considered a leading cause of cloud security incidents [5] and can often lead to unintended high costs (e.g., exposing an unencrypted S3 bucket resulting in massive data egress charges). AI-driven compliance tools can simultaneously check infrastructure for security best practices and cost efficiency. For example, a tool can flag publicly accessible S3 buckets (security risk) that also have an inappropriate storage class (cost inefficiency), demonstrating the convergence of security and financial governance in the operating model[13,18].

Conclusion on FinOps

The integration of advanced FinOps frameworks is now an operational necessity for any organization utilizing AWS at scale. The promise of the cloud—unlimited scalability and low OpEx—can only be realized through strict governance over the Inform, Optimize, and Operate cycles. By embracing automation and leveraging AI for predictive analysis, enterprises are positioned to ensure that the substantial economic benefits of migrating to AWS are not negated by the complexity of managing consumption, thereby helping to sustain the competitive advantage gained from cloud agility and elasticity [1, 2]. This sophisticated, interdisciplinary approach is what often separates merely "using" the cloud from effectively "mastering" it [16,17].

3.3. Security and Compliance Posture [13][18]

The security of the AWS cloud is governed by the Shared Responsibility Model, which is a crucial concept for any adopting organization [5]. In this model: [13][18]

- AWS is responsible for the Security of the Cloud: This includes the physical security of data centers, the infrastructure, networking, and the underlying hypervisor. [13][18]
- The Customer is responsible for Security in the Cloud: This covers customer data, operating systems (for IaaS services like EC2), application code, network configuration (Security Groups, NACLs), and, most critically, Identity and Access

Management (IAM) [5]. [13][18]

The availability of robust compliance certifications, supported by the platform's architecture, is associated with organizations meeting regulatory requirements (like HIPAA, PCI DSS, GDPR) with greater ease than building compliance into an on-premise environment [5]. AWS is seen to provide the compliant infrastructure, but the customer must correctly configure their services to maintain that compliance. [10][11][12]

Persistent Security Challenges: Despite the platform's inherent security, the most common security failures in the cloud are typically due to customer misconfiguration [5, 6]. Mismanaged IAM policies, overly permissive security group rules, and failure to encrypt data are common vectors. This underscores the human element in cloud security; the complexity of the platform places a high premium on specialized technical expertise[13,18].

4. Discussion and Conclusion

4.1. Synthesis of Findings and Theoretical Implications

The systematic analysis of AWS reveals a powerful synergy between its global architectural design and its disruptive economic model. The distributed nature of Regions and Availability Zones provides the technical foundation for unprecedented resilience and scalability, which, in turn, allows for the economic benefits of resource elasticity and consumption-based billing [3, 4]. The deep service ecosystem is associated with fostering modern, agile methodologies like DevOps [1, 2], confirming that the cloud is not merely a hosted data center but a complete, transformative operational platform.

The core theoretical implication is the successful commodification of IT infrastructure. By abstracting the complexities of physical hardware, power, and cooling, AWS has shifted the focus of IT departments from maintenance to innovation, a change that permeates every industry [6].

4.2. Strategic Implications for Enterprise Adoption

While the benefits are clear, strategic adoption of AWS is often associated with addressing the risks, primarily that of vendor lock-in. As enterprises integrate specialized, proprietary AWS services (like Amazon DynamoDB or various AI/ML services), the operational cost and complexity of migrating to another provider or back on-premise may increase dramatically. This is the trade-off for the deep integration and specialized functionality offered by the market leader [6]. Effective corporate strategy must therefore involve a rigorous assessment of multi-cloud or hybrid strategies, not necessarily to avoid AWS, but to help ensure critical, general-purpose workloads remain portable, mitigating concentration risk. Furthermore, with the growing need for ultra-low latency applications, the emergence of services like AWS Outposts (bringing AWS infrastructure to the customer's on-premise data center) suggests a new strategic trend toward Edge Computing that integrates the cloud's elasticity with local requirements [16,17].

4.3. Future Research Trajectories and Open Challenges

The findings of this review point to several crucial areas requiring future scholarly attention: [18][19]

1. Empirical Performance Validation: Further empirical studies are needed to provide independent, comparative performance benchmarks of serverless compute (Lambda) versus traditional virtual machines (EC2) under various real-world load conditions [7, 9]. This is valuable for practitioners to make data-driven decisions on service selection[14,15]
2. Advanced FinOps Models: The complexity of AWS billing requires the development of more advanced, perhaps AI-driven, FinOps models that can dynamically recommend instance resizing, commitment purchasing, and service switching in real-time to maintain continuous cost optimization.
3. Sustainability and Environmental Impact: As hyperscale data centers consume vast amounts of energy, research into the environmental sustainability of AWS operations and the effectiveness of its green initiatives is an increasingly critical ethical and operational challenge [19].

4.4. Limitations of the Current Study

The conclusions drawn in this review are subject to certain limitations:

- Reliance on Selective Sources: The strict constraint of utilizing only the nine provided reference sources [1–9] suggests that the review could not incorporate the entirety of the vast, most

current research on specific, rapidly deployed AWS features. [19]

- Proprietary Visibility: The proprietary nature of AWS architecture limits full-scope technical validation. Performance and security claims often rely on official documentation [3] and must be interpreted within that context.
- Rapid Obsolescence: The cloud market evolves quickly. Any static review, including this one, inherently struggles to capture the absolute latest service offerings or pricing changes that may have occurred between the publication dates of the cited sources.

4.5. Conclusion

Amazon Web Services (AWS) represents the pinnacle of modern utility computing, built on a resilient global architecture that fundamentally redefines the economics of IT infrastructure. By shifting the financial burden from high CapEx to manageable OpEx [4], and by fostering the operational efficiency associated with DevOps [1, 2], AWS continues to drive global digital transformation. However, strategic success is associated with mastering the new complexities of the cloud, particularly in the areas of FinOps (including its advanced AI-driven frameworks) and the Shared Responsibility Model [5, 6]. Future research must continue to validate performance, refine cost models, and address the strategic trade-offs inherent in deep platform commitment.

References

1. Edureka, https://www.edureka.co/blog/devops-tutorial
2. AWS official website – https://aws.amazon.com
3. Akshay Kushwaha. Research Paper on AWS Cloud Infrastructure vs Traditional On-Premise. International Research Journal of Engineering and Technology (IRJET), Volume 07, Issue 01, Jan 2020.
4. Neha Kewate, Amruta Raut, Mohit Dubekar, Yuvraj Raut, & Prof. Ankush Patil. A Review on AWS - Cloud Computing Technology. International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10, Issue I, Jan 2022.
5. Tec Target website – https://www.techtarget.com
6. https://www.irjet.net/archives/V7/i1/IRJET-V7I131.pdf
7. Figures and survey results (1 and 2): https://www.statista.com/statistics/511518/worldwide-survey-private-coud-services-running-application/
8. Figures and survey results (6 and 7)
9. Koneru, N. M. K. (2025). Containerization best practices: Using Docker and Kubernetes for enterprise applications. Journal of Information Systems Engineering and Management, 10(45s), 724–743. https://doi.org/10.55278/jisem.2025.10.45s.724
10. Murali Krishna Koneru, N. (2025). Centralized logging and observability in AWS: Implementing ELK Stack for enterprise applications. International Journal of Computational and Experimental Science and Engineering, 11(2). https://doi.org/10.22399/ijcesen.2289
11. Naga Murali Krishna Koneru. (2025). Leveraging AWS CloudWatch, Nagios, and Splunk for real-time cloud observability. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3781
12. Hariharan, R. (2025). Zero trust security in multi-tenant cloud environments. Journal of Information Systems Engineering and Management, 10(45s). https://doi.org/10.52783/jisem.v10i45s.8899
13. Reddy Dhanagari, M. (2025). Aerospike: The key to high-performance real-time data processing. Journal of Information Systems Engineering and Management, 10(45s), 513–531. https://doi.org/10.55278/jisem.2025.10.45s.513
14. Gannavarapu, P. (2025). Performance optimization of hybrid Azure AD join across multi-forest deployments. Journal of Information Systems Engineering and Management, 10(45s), e575–e593. https://doi.org/10.55278/jisem.2025.10.45s.575
15. Venkiteela, P. (2025). Modernizing opportunity-to-order workflows through SAP BTP integration architecture. International Journal of Applied Mathematics, 38(3s), 208–

228. https://doi.org/10.58298/ijam.2025.38.3s.12

16. Toffetti, G., Brunner, S., Blöchlinger, M., Spillner, J., & Bohnert, T. M. (2017). Self-managing cloud-native applications: Design, implementation, and experience. Future Generation Computer Systems, 72, 165–179.
17. Chadha, K. S. (2025). Zero-trust data architecture for multi-hospital research: HIPAA-compliant unification of EHRs, wearable streams, and clinical trial analytics. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3477
18. Kawaljeet Singh Chadha. (2025). Edge AI for real-time ICU alarm fatigue reduction: Federated anomaly detection on wearable streams. Utilitas Mathematica, 122(2), 291–308. https://utilitasmathematica.com/index.php/Index/article/view/2708
19. Prassanna Rao Rajgopal . Secure Enterprise Browser - A Strategic Imperative for Modern Enterprises. International Journal of Computer Applications. 187, 33 (Aug 2025), 53-66. DOI=10.5120/ijca2025925611
20. Kumar Tiwari, S. (2023). Integration of AI and machine learning with automation testing in digital transformation. International Journal of Applied Engineering & Technology, 5(S1), 95–103. Roman Science Publications.