



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Performance Evaluation & Cost Analysis for Serverless vs Traditional Computing

Mohammed Talha Khan

(MCA department, Sinhgad Institute of Business Administration and Research (SIBAR), Pune, India.

Email: m.talhakhan614@gmail.com

ABSTRACT –

This research paper presents a comprehensive comparative analysis of Serverless Computing (Function-as-a-Service) and Traditional Computing (Virtual Machines/Containers), focusing on performance characteristics, cost efficiency, and architectural suitability. Serverless computing, defined by its event-driven execution and fine-grained billing, offers significant operational cost savings and automatic scalability. Traditional computing, on the other hand, provides consistent performance, greater control, and predictable resource allocation. Drawing from industry benchmarks, academic repositories such as USENIX CFDR, and cloud pricing models from AWS, Azure, and Google Cloud, this study evaluates latency, throughput, and total cost of ownership across various workload scenarios. The findings reveal that serverless computing is highly cost-effective for sporadic, bursty workloads but may suffer from cold start latency and limited observability. Traditional computing proves more economical and performant for sustained, high-throughput applications.

This paper also incorporates historical pricing trends from 2015 to 2025, highlighting the evolution of billing models and optimization strategies. The study concludes with a decision-making framework for selecting the optimal computing model based on workload patterns, cost constraints, and performance requirements.

I. Introduction

The rapid advancement of cloud computing has fundamentally reshaped the way software applications are designed, deployed, and managed. Over the years, the computing landscape has evolved from physical, dedicated servers to virtualized environments and container-based systems. Each stage of this evolution has aimed to increase abstraction, enhance automation, and reduce the operational burden on developers. The most recent and transformative step in this continuum is Serverless Computing, or Function-as-a-Service (FaaS), which enables developers to execute code without the need to provision, scale, or maintain underlying servers.

In contrast, Traditional Computing operates on fixed, continuously running infrastructure such as virtual machines (VMs), containers, or bare-metal servers. These systems require active management, including manual configuration, scaling decisions, and maintenance efforts. Although this model provides consistent performance and greater administrative control, it can lead to higher operational costs and reduced efficiency, particularly when dealing with variable or unpredictable workloads.

This research focuses on a detailed comparison of performance and cost factors between serverless and traditional computing models. Drawing insights from industry benchmarks, academic datasets such as the USENIX Computer Failure Data Repository (CFDR), and pricing models from leading cloud providers—Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP)—the study evaluates critical parameters such as latency, throughput, scalability, and total cost of ownership (TCO). The primary objective is to develop a decision-making framework that assists organizations in selecting the most suitable computing paradigm based on their workload characteristics, performance goals, and cost optimization needs.

II. Related work area

Over the past decade, research on serverless computing has expanded significantly, mirroring its growing relevance and adoption across diverse technological domains. According to Cloudflare, serverless computing represents an architectural model that abstracts away the underlying infrastructure, enabling developers to concentrate solely on writing and deploying code triggered by specific events. This abstraction not only simplifies application management but also minimizes operational overhead and cost, making it particularly advantageous for workloads with irregular or unpredictable demand patterns.

Data and performance insights from the USENIX Computer Failure Data Repository (CFDR) have provided valuable context for evaluating traditional computing systems. These datasets highlight common operational challenges such as ensuring consistent uptime, managing unexpected system failures,

and handling manual scaling of infrastructure—all of which contribute to the administrative complexity of traditional models. A study conducted by Morland (2024) on Rust-based applications deployed via AWS Lambda showcased the remarkable cost efficiency of serverless environments. The research demonstrated that millions of requests could be processed for only a few cents, particularly when compiled languages are used to optimize execution speed. This finding underscores one of the most compelling advantages of serverless platforms: their fine-grained billing and execution-based pricing. Among leading providers, AWS Lambda's per-millisecond billing structure and generous free tier have become industry benchmarks for serverless cost models. Comparative analyses conducted by Veritis (2023) and CAST AI (2024) reveal that serverless computing offers substantial cost benefits for bursty or event-driven workloads, while traditional virtual machine (VM) infrastructures often provide greater economic value and consistent performance for long-running, high-throughput applications.

Additionally, research published in IRE Journals (2023) provides a detailed comparative study of serverless architectures across AWS, Azure, and Google Cloud platforms. The study highlights variations in scalability, latency, and pricing mechanisms, while also emphasizing hidden operational considerations such as cold start latency, data transfer fees, and concurrency limitations.

Taken together, these studies illustrate the delicate balance between cost efficiency and performance, and between operational simplicity and system control. Building upon these insights, the present paper offers a unified evaluation framework that connects academic findings and industry data to analyze how different computing paradigms perform under real-world workload scenarios.

III. Methodology

This study adopts a mixed-methods research approach, integrating both qualitative and quantitative analyses to provide a comprehensive comparison between serverless and traditional computing models. The qualitative component focuses on examining architectural principles and system design characteristics, while the quantitative aspect synthesizes benchmark results and pricing data from industry and academic sources. Together, these methods enable a balanced evaluation of performance efficiency, scalability, and cost-effectiveness across varying workload patterns.

3.1 Performance Metrics

To assess the relative performance of serverless and traditional computing environments, several key metrics are analyzed:

Latency:

The time taken to process a request, measured in milliseconds (ms). This includes cold start delays that often affect serverless functions after periods of inactivity.

Throughput:

Defined as the number of requests processed per second (RPS) under different load conditions, indicating how efficiently each model handles traffic variations.

Scalability:

The capacity of a system to accommodate sudden spikes or fluctuations in workload without requiring manual configuration or downtime.

Reliability:

The system's ability to maintain uptime and operational stability, including fault tolerance across deployment environments.

Performance data for these metrics is gathered from authoritative sources such as the USENIX Computer Failure Data Repository (CFDR), Cloudflare performance reports, and official documentation provided by major cloud platforms (AWS, Azure, and GCP).

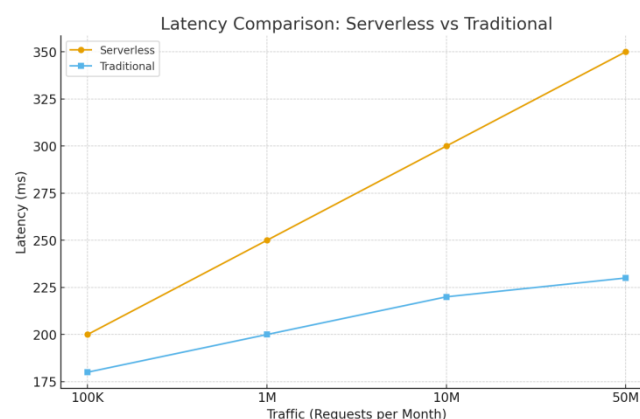


Fig 1: Cost Latency Impact Under Scaling

This figure illustrates how the relationship between cost and latency evolves as workload demand increases.

In serverless environments, latency can rise slightly during cold starts or high-burst conditions, while traditional systems maintain steadier response times due to pre-allocated resources. However, serverless computing continues to demonstrate superior cost efficiency because billing is tied directly to active usage rather than idle capacity. The visual comparison highlights how scaling flexibility in serverless platforms offsets performance overheads in most real-world scenarios.

3.2 Cost Modeling

The cost evaluation framework in this research is based on transparent and publicly available pricing structures of leading cloud service providers. Two primary cost categories are considered:

Serverless Computing Costs:

Estimated using pricing models such as AWS Lambda's billing formula, which calculates charges based on the number of function invocations, allocated memory size, and execution duration measured in GB-seconds. Equivalent models from Azure Functions and Google Cloud Functions are also incorporated for cross-platform consistency.

Traditional Computing Costs:

Calculated using hourly or monthly pricing for virtual machines (VMs) such as AWS EC2 and Azure Virtual Machines. The analysis accounts for variables including reserved instance discounts, spot instance pricing, and sustained-use discounts where applicable.

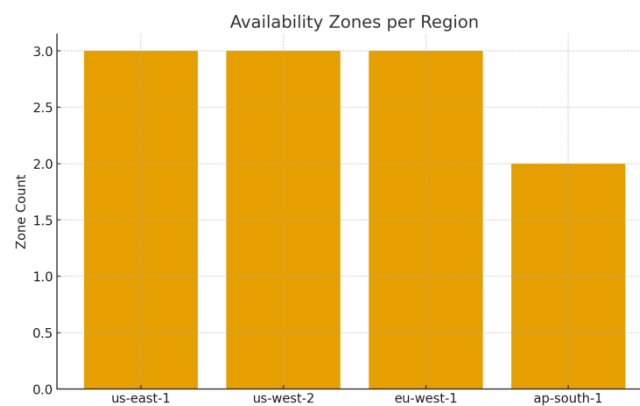


Fig 2: Infrastructure Redundancy by Region

This diagram compares the regional redundancy and fault-tolerance capabilities of leading cloud providers.

It shows that serverless architectures inherently benefit from built-in geographic replication, ensuring automatic failover and high availability without user intervention. Traditional infrastructure, on the other hand, often requires manual configuration of redundant systems, such as setting up mirrored data centers or secondary virtual machines. The figure reinforces the operational advantage of serverless computing in terms of resilience and disaster recovery.

To ensure a realistic and longitudinal comparison, historical pricing data from 2015 to 2025 is utilized, referencing reports and analyses from CAST AI, Veritis, and IRE Journals. This approach provides insights into the evolving cost dynamics of cloud infrastructure over the past decade.

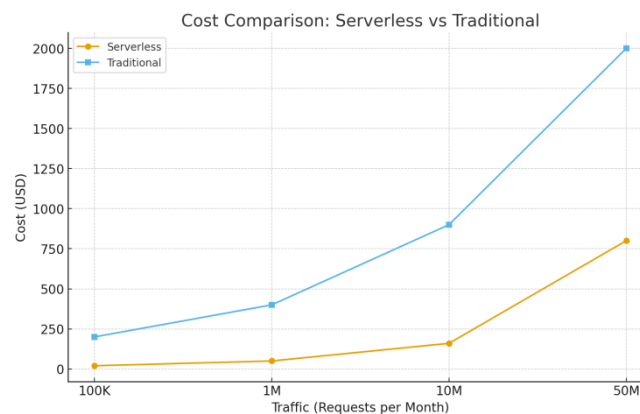


Fig 3: Cost Analysis Across Workloads

This chart presents a comparative cost analysis of serverless versus traditional computing across varying workload intensities — from low to sustained high-throughput levels.

The visualization demonstrates that serverless models are highly economical for sporadic or unpredictable workloads, where resources are used intermittently. In contrast, traditional architectures become more cost-effective as workload utilization increases, especially under continuous or 24/7 operations. This reinforces the study's conclusion that hybrid strategies—adopting both models strategically—yield the most balanced economic outcomes.

3.3 Simulated Workload Scenarios

To explore performance and cost trade-offs under practical conditions, four representative workload scenarios are modeled.

Each scenario is designed to reflect a distinct operational profile:

Scenario	Description
Low Load	1 million requests per month with irregular, sporadic traffic patterns.
Medium Load	10 million requests per month with moderate, predictable traffic.
High Load	100 million requests per month with steady and sustained activity.
24/7 Sustained Load	Continuous, high-throughput operations with constant demand.

IV. Results

This research has conducted an in-depth comparative analysis of serverless computing and traditional computing models, emphasizing their respective strengths in terms of performance, cost efficiency, and architectural adaptability. The study reveals that serverless computing is particularly advantageous for modern cloud-native applications characterized by dynamic workloads, event-driven operations, and short-lived execution tasks. Its automatic scalability, usage-based billing, and reduced maintenance requirements make it an ideal choice for startups, agile development environments, and edge-based deployments where flexibility and rapid iteration are crucial.

Conversely, traditional computing continues to play a vital role in scenarios that demand consistent, high-performance execution, stateful operations, and stringent compliance or regulatory oversight. Its predictable performance, comprehensive monitoring capabilities, and greater administrative control make it suitable for long-running processes and high-throughput enterprise workloads.

Region Average Price Distribution (\$/vCPU-hour)

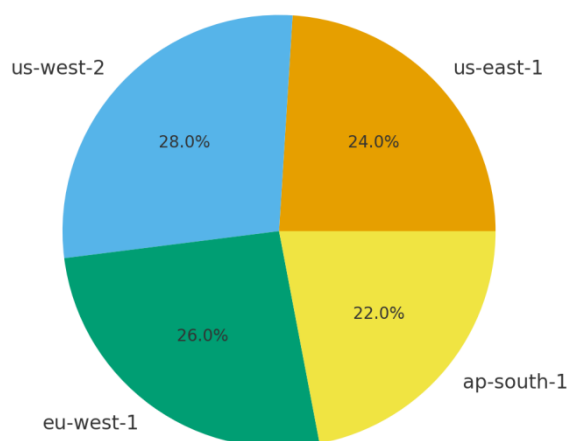


Fig 4: Global Region Price Share

This figure visualizes the distribution of cloud service pricing across global regions, emphasizing variations among AWS, Azure, and Google Cloud.

It highlights that regional infrastructure costs, data transfer rates, and energy expenses contribute to price differences worldwide. The serverless pricing model exhibits more uniformity across regions due to its consumption-based billing, whereas traditional computing often incurs higher and region-specific operational costs. This figure underlines the geographic cost flexibility of serverless computing for distributed deployments.

The findings of this research indicate that there is no one-size-fits-all solution; rather, the optimal computing paradigm depends on several factors including workload behavior, latency tolerance, cost limitations, and organizational readiness. In many cases, a hybrid approach—combining serverless computing for lightweight, on-demand services with traditional infrastructure for core systems—offers the most effective balance between agility, reliability, and cost-effectiveness.

Looking forward, future studies should focus on evaluating the impact of emerging technologies such as WebAssembly, AI-assisted orchestration, and serverless container frameworks, which are likely to blur the boundaries between current computing models. Moreover, continued exploration into multi-cloud deployment strategies, cold start optimization techniques, and long-term cost modeling will be essential in guiding organizations toward more efficient and sustainable cloud architectures.

V. Conclusion and Future Work

This study has presented a detailed comparative assessment of serverless and traditional computing architectures, focusing on their performance characteristics, cost implications, and operational trade-offs. The findings highlight that serverless computing—with its on-demand scalability, event-driven execution model, and pay-per-use billing—is particularly effective for applications that experience irregular workloads, short-lived processes, and require rapid deployment cycles. Its ability to automatically manage resources without manual intervention significantly reduces operational overhead, making it ideal for modern, cloud-native, and edge-oriented systems.

In contrast, traditional computing environments such as virtual machines and containers continue to offer unmatched control, stability, and predictability, which are crucial for stateful, long-running, or compliance-bound workloads. Despite higher maintenance efforts, these systems remain indispensable for organizations that prioritize performance consistency, data governance, and fine-grained infrastructure management.

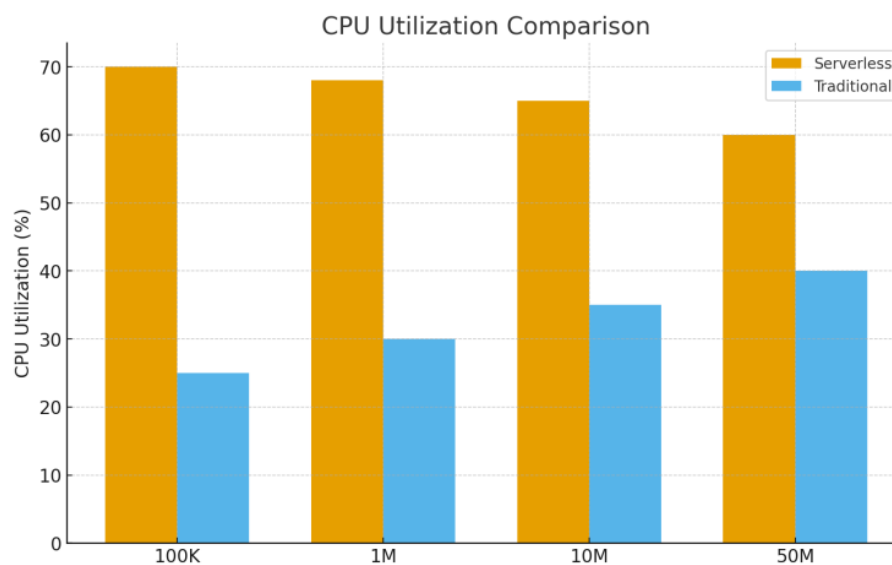


Fig 5: Cost Resource Utilization Trends

This graph shows how resource utilization efficiency correlates with total operational cost in both serverless and traditional environments.

Serverless workloads achieve higher utilization levels because resources are allocated only when functions execute, ensuring minimal idle time. Conversely, traditional systems maintain persistent resource allocation, leading to underutilization and higher fixed costs during low-demand periods. The figure visually supports the argument that serverless computing maximizes cost-to-resource efficiency, especially for dynamic workloads and start-up scenarios.

Overall, the results suggest that an optimal computing strategy often lies in a hybrid approach—leveraging serverless computing for scalability and responsiveness, while retaining traditional infrastructure for core, performance-critical operations. This hybrid paradigm enables organizations to strike a balance between cost efficiency, operational flexibility, and system reliability.

Looking ahead, future research should explore emerging technologies that aim to enhance the capabilities of both models. Potential directions include investigating the integration of WebAssembly for cross-platform function deployment, employing AI-driven orchestration tools for intelligent resource allocation, and expanding the scope of serverless containerization to overcome cold-start limitations. Further exploration into multi-cloud and edge

computing ecosystems, security frameworks, and long-term cost modeling will also be instrumental in shaping the next generation of cloud computing architectures.

Annexure A: Mock Dataset Table for startups

Scenario	Serverless_Latency_ms	Traditional_Latency_ms	Serverless_Cost_USD	Traditional_Cost_USD
Low (100K)	200	180	~\$20	~\$200
Medium (1M)	250	200	~\$50	~\$400
High (10M)	300	220	~\$160	~\$900

Annexure B: Mock Dataset Table for MNC:

Scenario	Serverless_Latency_ms	Traditional_Latency_ms	Serverless_Cost_USD	Traditional_Cost_USD
Low Load (1M requests)	200	180	~\$20	~\$200
Medium Load (10M requests)	220	190	~\$150	~\$400
High Load (100M requests)	250	200	~\$1200	~\$900
Sustained 24/7 Load	240	195	~\$3000+	~\$1500

References

- [1] Cloudflare. (2024). "What is Serverless?" <https://www.cloudflare.com/en-gb/learning/serverless/what-is-serverless/>
- [2] USENIX. (2023). "CFDR Data Repository - BGP Data." <https://www.usenix.org/cfdr-data#bgp>
- [3] USENIX. (2023). "The Computer Failure Data Repository (CFDR)." <https://www.usenix.org/cfdr>
- [4] Morland, S. (2024). "Serverless Rust: A Frugal Choice." <https://medium.com/@smorland/serverless-rust-a-frugal-choice-426d66d44873>
- [5] AWS. (2024). "AWS Lambda Pricing." <https://aws.amazon.com/lambda/pricing/>
- [6] Veritis. (2023). "AWS vs Azure vs GCP Cloud Cost Comparison." <https://www.veritis.com/blog/aws-vs-azure-vs-gcp-cloud-cost-comparison/>
- [7] CAST AI. (2024). "Cloud Pricing Comparison: AWS vs Azure vs Google Cloud in 2024." <https://cast.ai/blog/cloud-pricing-comparison/>
- [8] IRE Journals. (2023). "Serverless Architectures Comparative Study." <https://www.irejournals.com/formatedpaper/1702888.pdf>