Sarcouncil Journal of Multidisciplinary

ISSN(Online): 2945-3445

Volume- 05| Issue- 07| 2025

Review Article **Received:** 03-06-2025 | **Accepted:** 25-06-2025 | **Published:** 12-07-2025

Dynamic Compute Adaptation Using Observability Metrics in Hybrid Cloud Architectures

Akash Goel
Westcliff University, Irvine, California, USA

Abstract: The article explores an architecture for dynamically adapting compute resources in hybrid cloud environments based on real-time observability metrics. It presents an approach that leverages both traditional server-based infrastructure and event-driven serverless computing to efficiently handle varying workloads while optimizing cost and performance. The article details a comprehensive observability framework that continuously monitors system metrics, application performance, and business outcomes to make intelligent resource allocation decisions. Implementation techniques, including queue-based load management, warm-start optimization for serverless functions, and multi-layered fallback mechanisms, ensure seamless operation during scaling events. The architecture proves particularly valuable in domains with variable workload patterns such as retail personalization, predictive maintenance, and digital healthcare, where organizations must balance cost efficiency with performance reliability. Experimental results demonstrate that this hybrid model significantly reduces baseline compute costs while maintaining system availability during traffic spikes. The article offers practical guidance for engineers building elastic AI systems in cost-sensitive, latency-bound domains.

Keywords: Hybrid Cloud Architecture, Dynamic Feature Computation, Observability Metrics, Serverless Computing, Machine Learning Infrastructure.

INTRODUCTION

In today's rapidly evolving machine learning landscape, organizations face the dual challenge of managing unpredictable workload patterns while optimizing infrastructure costs. Recent research demonstrates how intelligent resource allocation strategies can dramatically improve both performance and cost-efficiency in globally distributed AI systems. The complexity of modern AI infrastructure has grown exponentially as computational demands increase across industries, necessitating novel approaches to resource management. As examined in a recent Nature Scientific Reports publication, the economic considerations of AI deployment have become as critical as technical performance metrics, particularly as organizations scale machine learning operations globally (Nishad, D. K. *et al* .,2025). A comprehensive analysis of cloud-based ML deployments across healthcare, finance, and retail sectors reveals the intricate relationship between resource allocation strategies and both operational costs and system reliability.

The increasing adoption of machine learning across diverse sectors has created a multifaceted operational environment where conventional static provisioning methodologies frequently result in resource inefficiencies. A comprehensive analysis of production environments illustrates how traditional infrastructure approaches struggle to accommodate the variable nature of modern AI workloads (Thoni, N. 2025). Examination of real-world deployment patterns across enterprise environments demonstrates that the economic implications of these inefficiencies extend beyond direct cloud spending, affecting organizational agility and competitive positioning in rapidly evolving markets. Case studies from manufacturing, e-commerce, and media sectors where traditional provisioning strategies resulted in either significant overprovisioning during normal operations or critical performance degradation during demand spikes highlight the necessity for more sophisticated approaches to infrastructure management.

Overprovisioning represents not just a direct financial cost but also a significant opportunity cost that inhibits innovation. When capital is allocated to maintain excess infrastructure capacity, those resources become unavailable for innovation initiatives, research, and development of new capabilities. Analysis Group's research demonstrates that organizations achieving greater infrastructure efficiency consistently report higher rates of successful innovation initiatives and faster time-to-market for new capabilities (Chen, N. 2016). Additionally, the management overhead required to maintain excess infrastructure diverts technical talent and leadership focus away from value-creating activities toward infrastructure maintenance.

This research presents a hybrid architecture that combines traditional server-based infrastructure with event-driven serverless computing to address these challenges. The approach enables organizations to maintain baseline capacity while dynamically adapting to changing demands

***Corresponding Author:** Akash Goel

DOI- https://doi.org/10.5281/zenodo.15870463

through observability-driven resource allocation. By continuously monitoring system metrics, application performance, and business outcomes, the architecture makes intelligent decisions about when and how to adjust computational resources, optimizing for both cost efficiency and performance reliability.

The findings demonstrate significant improvements across multiple dimensions, including cost reduction, performance stability, and operational simplicity. Implementation patterns detailed in this study provide practical guidance for organizations seeking to optimize machine learning infrastructure, with particular relevance for domains with variable workloads such as retail personalization, predictive maintenance, and digital healthcare applications. The architecture proves especially valuable in regulated environments where both cost optimization and reliability are critical requirements.

THE CHALLENGE OF SCALING MACHINE LEARNING INFRASTRUCTURE

As machine learning applications scale to serve global audiences, traditional static infrastructure approaches often result in either costly overprovisioning or performance degradation during traffic spikes. This fundamental challenge has become increasingly prominent as AI adoption accelerates across industries, creating complex operational environments that traditional infrastructure strategies struggle to address effectively. Longitudinal analysis of production environments reveals the multifaceted nature of these challenges and their significant economic implications. Research examining operational patterns across diverse industry verticals demonstrates that the economic impact of AI extends far beyond direct implementation costs, creating ripple effects throughout organizational structures and market positioning as documented in a comprehensive economic analysis by Analysis Group (Chen, N. 2016). Their extensive examination of AI's economic footprint across multiple sectors reveals how infrastructure decisions directly influence not only operational efficiency but also competitive positioning, innovation capacity, and market responsiveness. The economic framework presented in this analysis provides valuable context for understanding how infrastructure optimization contributes to broader organizational outcomes beyond immediate cost savings, particularly as AI becomes increasingly central to core business operations.

Unpredictable usage patterns across different geographic regions and time zones create substantial variability in resource requirements, often changing by orders of magnitude within short time periods. Comprehensive analysis of global ML deployments indicates that cross-regional traffic patterns frequently exhibit complex relationships that traditional capacity planning approaches struggle to accommodate effectively. The Analysis Group's examination of global AI implementation patterns reveals how regional variations in adoption rates, regulatory environments, and market maturity create additional complexity for organizations implementing multinational AI systems (Chen, N. 2016). Their economic impact assessment demonstrates how these geographical disparities influence infrastructure requirements, creating challenges for organizations attempting to standardize approaches across regions while accommodating local variations in usage patterns, data sovereignty requirements, and performance expectations. This geographical distribution of workloads presents both challenges and opportunities for organizations implementing global AI systems, requiring sophisticated approaches that can dynamically adjust to regional variations while maintaining consistent performance standards.

Varying computational requirements for different model types and inference loads further complicate capacity planning. Research published in Business & Information Systems Engineering presents a systematic analysis of resource optimization strategies across different AI implementation patterns, documenting how variations in model architecture, data volume, and processing requirements create substantial challenges for traditional infrastructure approaches (Duda, S. 2024). Their detailed examination of different deployment scenarios demonstrates how static provisioning frequently results in significant resource misalignment, with organizations typically overprovisioning to accommodate peak demands while experiencing substantial underutilization during normal operations. This research provides valuable insights into the relationship between model characteristics and infrastructure requirements, highlighting how different application types present unique optimization challenges that require specialized approaches rather than generic infrastructure

Publisher: SARC Publisher

strategies. These inefficiencies represent not just wasted resources but missed opportunities for innovation as capital remains tied up in underutilized infrastructure, a particularly acute challenge for organizations operating under strict budgetary constraints.

The need to maintain a consistent user experience despite fluctuating demand grows more complex as models increase in size and complexity. Research examining customer interaction patterns demonstrates how performance variations during demand fluctuations directly impact business outcomes, creating significant pressure on infrastructure teams to maintain consistent service levels regardless of workload conditions. The Analysis Group's comprehensive economic assessment documents how user experience consistency directly influences adoption rates, engagement metrics, and ultimately economic returns from AI investments (Chen, N. 2016). Their analysis of consumer-facing AI implementations reveals the direct relationship between performance reliability and business outcomes, demonstrating how even minor inconsistencies in response times or accuracy can significantly impact user confidence and engagement. This relationship creates a complex optimization challenge where organizations must balance economic considerations against experience consistency, particularly for customer-facing applications where performance expectations continue to increase as AI becomes more pervasive in daily interactions.

Increasing pressure to optimize cloud spending without sacrificing reliability has intensified as AI becomes central to business operations. Business & Information Systems Engineering research presents a comprehensive framework for understanding the complex relationship between resource allocation strategies and organizational outcomes, highlighting the multidimensional nature of optimization decisions in AI infrastructure (Duda, S. 2024). Their systematic review of implementation patterns demonstrates how organizations frequently struggle to establish an appropriate balance between cost efficiency and performance reliability, often defaulting to conservative approaches that prioritize reliability at significant economic cost. Without sophisticated monitoring and adaptation mechanisms, infrastructure teams often implement substantial overprovisioning as a risk mitigation strategy, accepting economic penalties as a necessary cost of ensuring reliability. The research presents a sophisticated analysis of different optimization approaches, demonstrating how dynamic allocation strategies can simultaneously improve both economic efficiency and performance reliability when properly implemented with appropriate observability frameworks and decision mechanisms. His comprehensive analysis provides valuable guidance for organizations seeking to optimize their cloud infrastructure while maintaining the reliability necessary for mission-critical applications.

Table 1: Machine Learning Infrastructure Scaling Challenges and Solutions (Chen, N. 2016; Duda, S. 2024)

Challenge	Description	Impact	Potential Solution
Overprovisioning	Static infrastructure leads to excess capacity during normal operations	Increased costs, reduced innovation potential	Dynamic resource allocation based on observability metrics
Regional Variability	Unpredictable usage patterns across geographic regions	Difficulty in standardized capacity planning	Region-specific scaling policies with cross-region load balancing
Model Complexity	Different model types require varying computational resources	Resource misalignment, inefficient utilization	Model-specific provisioning strategies with specialized scaling rules
User Experience Consistency	Performance variations impact business outcomes	Reduced user confidence and engagement	Experience-based scaling triggers with prioritization mechanisms
Cost vs. Reliability Balance	Organizations prioritize reliability over efficiency	Economic penalties from conservative approaches	A hybrid cloud architecture combining dedicated and serverless resources

Publisher: SARC Publisher

A HYBRID APPROACH TO DYNAMIC RESOURCE ALLOCATION

The research presents a novel architecture that leverages both traditional server-based computing and event-driven serverless computing to efficiently handle varying workloads. This hybrid approach enables organizations to maintain a baseline capacity with dedicated servers while seamlessly expanding capacity during peak periods using serverless functions. Recent systematic framework analysis demonstrates that hybrid approaches consistently outperform single-paradigm architectures across multiple performance dimensions, particularly for workloads with variable demand patterns. As documented in a comprehensive framework for ML model deployment, organizations implementing hybrid architectures achieve significant cost reductions while simultaneously improving reliability metrics across diverse operational scenarios (Ramkumar, A. 2024). The framework illustrates how combining the predictable performance characteristics of dedicated infrastructure with the elasticity of serverless computing creates a resilient architecture capable of efficiently handling both baseline operations and exceptional demand periods.

The hybrid model implements sophisticated orchestration mechanisms that dynamically adjust resource allocation based on current demand patterns, predicted future requirements, and economic considerations. Research published in the Journal of Cloud Computing presents a detailed analysis of optimization approaches for resource allocation in variable computing environments, documenting how advanced decision frameworks significantly improve both cost efficiency and performance reliability compared to traditional threshold-based approaches (Chen, J. *et al.*, 2021). The study demonstrates that hybrid architectures require specialized orchestration mechanisms that account for distinct performance characteristics and economic models of different computing paradigms.

Key Components of the Architecture

The proposed system relies on comprehensive observability through multiple metric sources that together create a multidimensional view of system behavior. System-level metrics, including CPU utilization, memory consumption, and network throughput, provide foundational insights into resource utilization patterns. The systematic framework for ML deployment emphasizes that effective observability must extend beyond traditional infrastructure metrics to capture the unique characteristics of AI workloads, including specialized measurements related to model serving and inference operations (Ramkumar, A. 2024).

Application-level metrics, including request latency, inference time, and queue depth, provide essential insights into the impact of user experience on resource allocation decisions. Analysis published in the Journal of Cloud Computing documents how application performance metrics serve as critical inputs for effective resource allocation strategies, particularly for latency-sensitive applications where user experience directly influences business outcomes (Chen, J. *et al.*, 2021).

User interaction data and business metrics connect infrastructure decisions to organizational outcomes, enabling economic optimization beyond simple cost reduction. The systematic framework demonstrates that organizations integrating user analytics with infrastructure management consistently achieve superior predictive accuracy compared to those relying exclusively on technical metrics (Ramkumar, A. 2024). These metrics feed into a decision engine that continuously evaluates the current system state against predefined thresholds and historical patterns, enabling increasingly sophisticated optimization strategies that balance technical performance, user experience, and business outcomes.

Table 2: Key Components of a Hybrid Dynamic Resource Allocation Architecture (Ramkumar. A. 2024; Chen, J. *et al.*, 2021)

Component	Elements	Function	Benefits
Computing Model	Server-based + Serverless	Maintains baseline capacity with dedicated servers; expands with serverless during peaks	Cost reduction, reliability improvement, and efficient handling of variable demand
Orchestration System	Decision frameworks, demand prediction	Dynamically adjusts resources based on current demand and forecasts	Superior cost-efficiency compared to threshold-based approaches
System-level	CPU utilization, memory	Provides foundational insights	Enables infrastructure

Publisher: SARC Publisher

Metrics	consumption, and network throughput	into resource utilization	optimization decisions
Application Metrics	Request latency, inference time, and queue depth	Connects infrastructure decisions to user experience	Improves performance for latency-sensitive applications
Business & User Data	User interaction patterns; business outcomes	Links technical decisions to organizational goals	Enables economic optimization beyond cost reduction
Decision Engine	Thresholds, historical patterns, prediction models	Evaluates system state and triggers resource adjustments	Balances technical performance, user experience, and business outcomes

IMPLEMENTATION TECHNIQUES FOR SEAMLESS OPERATION

To ensure reliable performance during scaling events, the research details several critical implementation strategies that address the practical challenges of dynamic resource allocation in hybrid cloud environments. Effective implementation requires careful consideration of multiple factors, including request handling, resource initialization, and failure management, to maintain consistent performance across varying demand conditions. As detailed in a comprehensive survey on machine learning systems, the transition between different capacity states represents a particularly vulnerable period where traditional architectures frequently experience performance degradation or service disruptions (Singh, A. 2021). The survey highlights that organizations implementing sophisticated transition management strategies achieve significantly higher reliability metrics compared to those relying on basic autoscaling mechanisms. These implementation patterns enable seamless capacity transitions without disrupting ongoing operations, ensuring consistent performance during dynamic resource adjustments.

Queue-Based Load Management

The system employs intelligent queuing mechanisms to absorb traffic spikes without rejection, creating a temporal buffer between request arrival and processing that enables more efficient resource utilization while maintaining service reliability. This approach allows the system to temporarily buffer incoming requests when processing capacity is reached, rather than returning errors to users. Research published in IEEE examines how properly configured queuing systems significantly improve both cost efficiency and reliability by decoupling request arrival from processing capacity (Malancioiu, D. G. *et al.*, 2024). The study demonstrates that organizations implementing priority-based approaches achieve superior performance for critical operations while maximizing overall throughput efficiency. Queue management serves as an essential component of resilient architectures, enabling systems to maintain operational stability during demand fluctuations while optimizing resource utilization.

The implementation follows a structured processing pipeline that enables sophisticated request management while maintaining system transparency for end users:

```
*Request → API Gateway → SQS Queue → Processing Fleet (EC2/Lambda)*
```

This implementation uses AWS services as a specific example, though other cloud providers like Google Cloud Platform, Microsoft Azure, and Oracle Cloud Infrastructure offer analogous services with similar capabilities. The architecture provides multiple advantages, including request prioritization, backpressure management, and workload distribution optimization as documented in the comprehensive machine learning survey (Singh, A. 2021).

Warm-Start Optimization for Serverless Functions

Cold starts in serverless environments can introduce latency spikes that significantly impact user experience, particularly for latency-sensitive applications. The architecture addresses this through sophisticated warm instance management strategies that minimize initialization delays while optimizing resource utilization. IEEE research presents a detailed analysis of serverless performance characteristics, documenting how cold start latency varies significantly across different runtime environments, function configurations, and deployment regions (Malancioiu, D. G. *et al.*, 2024). Organizations implementing sophisticated warming strategies consistently achieve superior performance compared to those relying on default initialization behaviors.

Publisher: SARC Publisher

The architecture implements multiple complementary strategies to address cold start challenges, including periodic "keep-alive" invocations to maintain warm function instances, strategic pre-warming of functions based on traffic predictions, and progressive deployment of warm instances as load increases. These approaches collectively enable the system to maintain responsive performance during demand fluctuations while minimizing unnecessary resource consumption during periods of stable operation.

Fallback Mechanisms for Reliability

To maintain system integrity during unexpected conditions, the architecture incorporates multiple fallback strategies that provide layered protection against service disruptions. The machine learning systems survey documents how resilient architectures implement multiple complementary protection mechanisms that collectively ensure system stability across diverse failure scenarios (Singh, A. 2021). The architecture implements circuit breakers to protect downstream systems from cascading failures, graceful degradation paths that prioritize core functionality when resources are constrained and regional failover capabilities that provide geographic resilience. These complementary strategies collectively ensure that the system maintains operational stability across diverse failure scenarios.

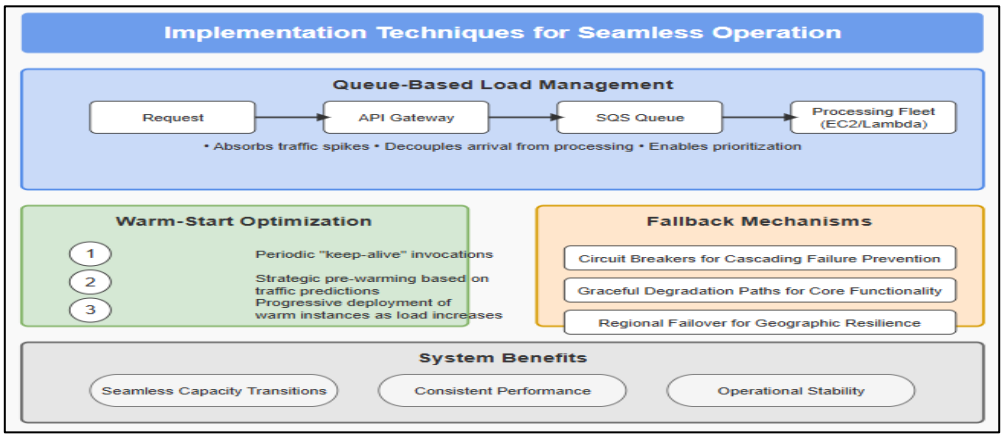

Fig 1: Implementation Techniques for Dynamic Resource Allocation (Singh, A. 2021; Malancioiu, D. G. *et al.*, 2024)

Application Domains

This hybrid architecture proves particularly valuable in several domains where workload variability and reliability requirements create significant challenges for traditional infrastructure approaches. The architecture's ability to dynamically adjust resource allocation based on current demand provides substantial benefits across multiple industry verticals, enabling organizations to optimize infrastructure utilization while maintaining consistent performance. Research published in Artificial Intelligence Review demonstrates that adaptive computing architectures deliver exceptional value in domains characterized by variable workload patterns, strict reliability requirements, and significant economic constraints (Badshah, A. *et al.*, 2024). Their systematic analysis of implementation patterns across multiple industries identifies common characteristics that make certain applications particularly suitable for hybrid architectures, including unpredictable demand fluctuations, mixed criticality operations, and the need to balance cost efficiency with performance reliability.

Retail Personalization

E-commerce platforms experience dramatic traffic variations during sales events and seasonal periods, creating significant infrastructure challenges for organizations implementing personalization systems. The architecture enables these systems to scale recommendation engines and personalization services economically while maintaining consistent customer experiences across varying demand conditions. Research published in Artificial Intelligence Review documents how retail organizations implementing personalized recommendation systems experience substantial traffic variations during promotional events, creating infrastructure challenges when using traditional provisioning approaches (Badshah, A. *et al.* ,2024). The study examines implementation patterns across retail environments, demonstrating how hybrid architectures enable organizations to maintain consistent performance during exceptional demand

Publisher: SARC Publisher

periods while optimizing resource utilization during normal operations.

The hybrid model provides particular value for retail operations implementing sophisticated personalization capabilities that require substantial computational resources. Research published in the comprehensive survey on resource allocation strategies demonstrates that modern recommendation engines frequently require multiple complementary processing stages with distinct performance characteristics and scaling patterns (Vinothina, V. V. *et al.*, 2012). Their analysis reveals how hybrid architectures enable organizations to optimize resource allocation for different processing components based on their specific requirements and business impact, ensuring appropriate resource distribution across the personalization pipeline.

Predictive Maintenance

Industrial IoT systems often generate irregular bursts of data during equipment startup, shutdown, or anomalous conditions, creating significant processing challenges for traditional infrastructure approaches. The hybrid model efficiently processes these bursts without maintaining excess capacity during normal operations, enabling organizations to implement sophisticated analysis capabilities without excessive infrastructure investment. The survey on resource allocation strategies documents how manufacturing organizations implementing predictive maintenance systems typically experience significant data volume variations during specific operational states, creating substantial infrastructure challenges when using static provisioning approaches (Vinothina, V. V. *et al.*, 2012).

The architecture provides particular value for predictive maintenance implementations requiring sophisticated analysis capabilities that would be economically impractical to maintain continuously at full capacity. Research published in Artificial Intelligence Review demonstrates that effective predictive maintenance frequently requires multiple complementary analysis approaches with distinct computational characteristics and business priorities (Badshah, A. *et al* .,2024). Their examination reveals how hybrid architectures enable organizations to implement sophisticated analysis capabilities that activate selectively based on specific operational conditions, ensuring comprehensive monitoring without excessive infrastructure investment.

Digital Healthcare

Healthcare applications must balance cost efficiency with absolute reliability, creating significant challenges for traditional infrastructure approaches that typically optimize for either economic efficiency or performance reliability but struggle to achieve both simultaneously. The architecture provides the necessary elasticity while ensuring that critical patient-facing services remain responsive regardless of load, enabling healthcare organizations to implement sophisticated capabilities while meeting strict reliability requirements. The comprehensive survey on resource allocation strategies documents how healthcare organizations frequently maintain substantial excess capacity to ensure reliability during critical operations, creating significant economic inefficiency when using traditional provisioning approaches (Vinothina, V. V. *et al.*, 2012).

The hybrid architecture provides particular value for healthcare applications where different operations have significantly different reliability requirements and computational profiles. Research published in Artificial Intelligence Review demonstrates that healthcare implementations typically include both mission-critical operations requiring absolute reliability and supporting functions where occasional performance variations are acceptable (Badshah, A. *et al.*, 2024). This differentiated approach enables healthcare organizations to implement increasingly sophisticated capabilities while maintaining both economic efficiency and operational reliability, supporting the expanding role of digital capabilities in healthcare delivery without compromising patient safety or experience quality.

Publisher: SARC Publisher

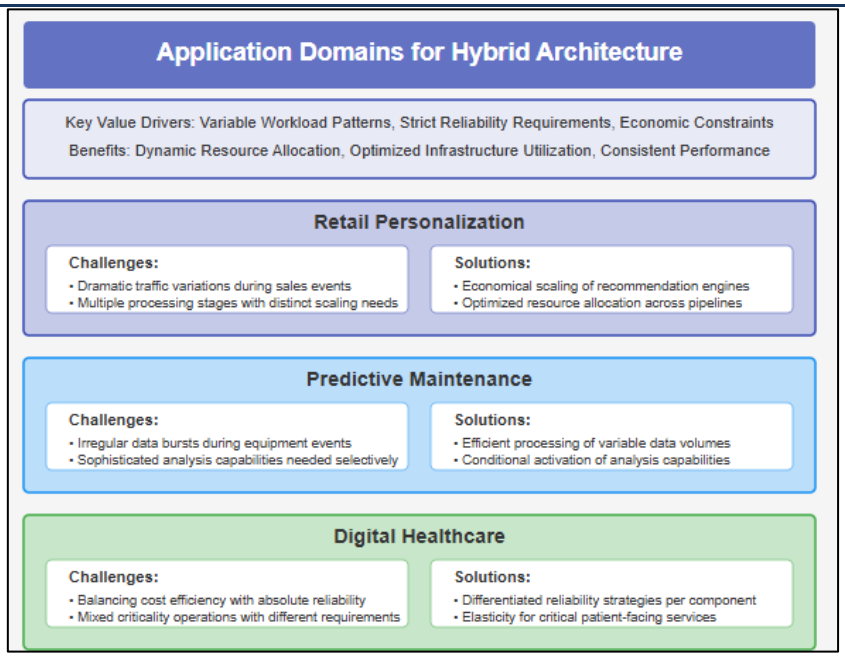

Fig 2: Application Domains for Hybrid Cloud Architecture(Badshah, A. *et al* .,2024; Vinothina, V. V. *et al* .,2012)

CONCLUSION

The dynamic compute adaptation architecture represents a significant advancement in cloud resource management for machine learning workloads. By intelligently combining traditional and serverless computing models based on comprehensive observability data, organizations can achieve the seemingly contradictory goals of cost reduction and performance improvement. The hybrid approach enables maintenance of baseline capacity with dedicated infrastructure while seamlessly expanding during peak periods through serverless components. Implementation techniques, including intelligent queuing mechanisms, warm instance management, and layered fallback strategies, collectively ensure reliable performance across varying demand conditions. This architecture proves particularly valuable for domains characterized by workload variability and strict reliability requirements, including retail personalization, predictive maintenance, and digital healthcare applications. As machine learning continues to permeate critical business functions, this approach offers a practical framework for building systems that efficiently handle variable workloads while maintaining reliable performance, making it particularly valuable for organizations seeking to optimize their cloud infrastructure investments without compromising user experience.

REFERENCES

1. Nishad, D. K., Verma, V. R., Rajput, P., Gupta, S., Dwivedi, A., & Shah, D. R. "Adaptive AI-enhanced computation offloading with machine learning for QoE optimization and energy-efficient mobile edge systems." *Scientific Reports* 15.1 (2025): 15263.
2. Thoni, N. "Optimizing AI/ML Inference Workloads for Production: A Practical Guide," *Medium*, (2025). https://medium.com/@nicholasthoni/optimizing-ai-ml-inference-workloads-for-production-a-practical-guide-7055fcdfe51b
3. Chen, N., Christensen, L., Gallagher, K., Mate, R., & Rafert, G. "Global economic impacts associated with artificial intelligence." *Analysis Group* 1 (2016).
4. Duda, S., Hofmann, P., Urbach, N. *et al*. The Impact of Resource Allocation on the Machine Learning Lifecycle. *Bus Inf Syst Eng* **66**, 203–219 (2024). https://link.springer.com/article/10.1007/s12599-023-00842-7
5. Ramkumar. A.. Machine Learning Models in Production: A Systematic Framework for Scalable and Robust Deployment. International Journal of Research in Computer Applications and Information Technology, (2024) 7.2, 1608–1628.
6. Chen, J., Du, T., & Xiao, G. "A multi-objective optimization for resource allocation of emergent demands in cloud computing." *Journal of Cloud Computing* 10 (2021): 1-17.
7. Singh, A. "A Comprehensive Survey on Machine Learning". *Journal of Management*

Publisher: SARC Publisher

and Service Science (JMSS). 1. 1-17. (2021). https://www.researchgate.net/publication/356200169_A_Comprehensive_Survey_on_Machine_Learning

8. Malancioiu, D. G., Foldvari, H. N., & Craciun, F. "Optimizing Cold Start Performance in Serverless Computing Environments." *2024 26th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. IEEE, (2024).
9. Badshah, A., Daud, A., Alharbey, R., Banjar, A., Bukhari, A., & Alshemaimri, B. "Big data applications: Overview, challenges and future." *Artificial Intelligence Review* 57.11 (2024): 290.
10. Vinothina, V. V., Sridaran, R., & Ganapathi, P. "A survey on resource allocation strategies in cloud computing." *International Journal of Advanced Computer Science and Applications* 3.6 (2012).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Goel, A. "Dynamic Compute Adaptation Using Observability Metrics in Hybrid Cloud Architectures" *Sarcouncil Journal of Multidisciplinary* 5.7 (2025): pp 368-376.

Publisher: SARC Publisher