IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

Review of Machine Learning-Based Scheduling Frameworks in Scalable Cloud Computing Architectures

Nitin Soni[1] and Dr. Nilesh Vasant Ingale[2]
[1]Research Scholar, Department of Computer Science
[2]Associate Professor, Department of Computer Science
Vikrant University, Gwalior M.P

Abstract: *Cloud computing has emerged as a dominant paradigm for delivering scalable computing resources over the internet. The rapid growth of big data applications, Internet of Things, artificial intelligence, and distributed enterprise systems has significantly increased the demand for efficient scheduling frameworks in cloud environments. Traditional scheduling techniques often fail to handle dynamic workloads, heterogeneous resources, and real-time processing requirements.*

Machine Learning-based scheduling frameworks have gained attention because of their capability to learn workload patterns, predict resource demands, and optimize scheduling decisions automatically. This review paper discusses various machine learning-based scheduling frameworks used in scalable cloud computing architectures. It highlights major ML techniques, scheduling objectives, benefits, challenges, and future research directions associated with intelligent cloud scheduling systems..

Keywords*:* Machine Learning, Cloud Computing, Task Scheduling, Resource Allocation, Scalable Architectures

I. INTRODUCTION

Cloud computing provides on-demand access to computational resources such as servers, storage, software, and networking services. The scalability and flexibility of cloud environments allow organizations to process large-scale applications efficiently. However, increasing workloads and diverse user requirements create challenges related to task scheduling, resource allocation, load balancing, and energy management. Scheduling is considered one of the most important functions in cloud computing because it determines how tasks are assigned to available resources for optimal performance.

Traditional scheduling algorithms such as First Come First Serve, Round Robin, Min-Min, and Max-Min operate using static rules and predefined assumptions. These methods are insufficient for handling highly dynamic cloud infrastructures. Machine learning-based scheduling frameworks provide adaptive and intelligent solutions capable of predicting workload behavior and optimizing resource utilization in real time.

Researchers have explored supervised learning, unsupervised learning, reinforcement learning, deep learning, and hybrid ML techniques for scheduling optimization. These intelligent approaches improve Quality of Service, reduce execution time, lower energy consumption, and increase scalability in distributed cloud architectures.

EVOLUTION OF CLOUD SCHEDULING TECHNIQUES

Initially, cloud scheduling relied on heuristic and rule-based approaches. These algorithms were designed to minimize task completion time and maximize CPU utilization. With the expansion of cloud services, scheduling problems became more complex because of virtualization, heterogeneous resources, and fluctuating workloads.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, January 2026 Impact Factor: 7.67

The integration of machine learning introduced self-adaptive scheduling systems capable of learning from historical execution data. ML-based schedulers analyze system behavior continuously and make predictive decisions regarding workload distribution and resource provisioning.

I. Traditional Scheduling Approaches
First Come First Serve
Round Robin Scheduling
Priority Scheduling
Min-Min and Max-Min Algorithms
Genetic Algorithms
Ant Colony Optimization

II. Intelligent Scheduling Approaches
Neural Network-Based Scheduling
Reinforcement Learning Scheduling
Deep Learning Optimization
Fuzzy Logic Scheduling
Hybrid Evolutionary ML Frameworks

MACHINE LEARNING TECHNIQUES USED IN CLOUD SCHEDULING

A. Supervised Learning
Supervised learning techniques use labeled training datasets to predict future scheduling behavior. Regression and classification algorithms help estimate resource requirements and execution times.
Applications include:
Workload prediction
VM allocation
Resource demand forecasting
Task classification
Common algorithms:
Decision Trees
Support Vector Machines
Random Forest
Linear Regression

B. Unsupervised Learning
Unsupervised learning identifies hidden patterns in cloud workload datasets without labeled outputs. Clustering algorithms group similar tasks and optimize scheduling efficiency.
Applications include:
Workload clustering
Anomaly detection
Resource grouping
Dynamic load balancing
Common algorithms:
K-Means Clustering
Hierarchical Clustering
Self-Organizing Maps

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

C. Reinforcement Learning

Reinforcement learning allows scheduling agents to learn optimal decisions through interaction with the cloud environment. RL-based schedulers improve performance over time using reward mechanisms.

Applications include:
Dynamic VM scheduling
Energy-aware scheduling
Autonomous resource management
Common methods:
Q-Learning
Deep Q Networks
Markov Decision Processes

D. Deep Learning

Deep learning models process large-scale cloud datasets and identify complex scheduling patterns. These models are highly effective for predictive analytics in scalable cloud systems.

Applications include:
Real-time workload forecasting
Intelligent resource provisioning
Traffic prediction
Common architectures:
Convolutional Neural Networks
Recurrent Neural Networks
Long Short-Term Memory

ARCHITECTURE OF ML-BASED SCHEDULING FRAMEWORKS

A typical machine learning-based scheduling framework consists of multiple interconnected components that monitor cloud activities and optimize resource allocation.

I. Main Components

Data Collection Layer
Feature Extraction Module
Machine Learning Engine
Scheduling Decision Manager
Resource Allocation Layer
Performance Monitoring System

The framework continuously collects execution logs, CPU usage patterns, network traffic data, and storage metrics. ML algorithms process this information to predict future workload demands and generate optimized scheduling decisions.

Table 1. Comparative Analysis of ML-Based Scheduling Techniques

ML Technique	Scheduling Objective	Advantages	Limitations
Decision Trees	Task Classification	Simple and interpretable	Limited scalability
Random Forest	Resource Prediction	High accuracy	Computational overhead
Reinforcement Learning	Dynamic Scheduling	Self-adaptive learning	Long training time
Deep Learning	Workload Forecasting	Handles complex data	High resource consumption
K-Means Clustering	Load Balancing	Efficient grouping	Sensitive to cluster size
Genetic Algorithms with ML	Multi-objective Optimization	Better global optimization	Slow convergence

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

Table 2. Performance Metrics in Cloud Scheduling

Metric	Description
Makespan	Total completion time of tasks
Throughput	Number of tasks processed per unit time
Resource Utilization	Efficiency of resource usage
Energy Consumption	Power usage during execution
Response Time	Delay between request and execution
SLA Violation Rate	Failure to meet service agreements
Scalability	Ability to handle workload growth

ADVANTAGES OF ML-BASED SCHEDULING FRAMEWORKS

Machine learning-based scheduling frameworks provide several advantages in cloud computing architectures:

1. Improved Resource Utilization

ML algorithms optimize resource allocation dynamically, reducing idle resources and improving computational efficiency.

2. Intelligent Workload Prediction

Predictive models forecast future workloads accurately, enabling proactive resource provisioning.

3. Reduced Energy Consumption

Energy-aware scheduling minimizes power usage in data centers by optimizing server utilization.

4. Enhanced Scalability

ML frameworks adapt effectively to large-scale cloud infrastructures and varying workloads.

5. Better Quality of Service

Intelligent scheduling reduces latency, execution time, and SLA violations.

6. Autonomous Decision Making

Reinforcement learning enables self-learning cloud systems capable of automated scheduling decisions.

CHALLENGES IN ML-BASED CLOUD SCHEDULING

1. Data Complexity

Cloud environments generate massive volumes of heterogeneous data that require advanced preprocessing techniques.

2. Computational Overhead

Complex ML models consume significant computational resources during training and inference.

3. Scalability Issues

Some ML algorithms face difficulties handling extremely large distributed systems.

4. Security and Privacy Concerns

Sensitive cloud data used for model training may create security vulnerabilities.

5. Real-Time Adaptation

Achieving low-latency decision making in real-time cloud environments remains challenging.

EMERGING TRENDS IN INTELLIGENT CLOUD SCHEDULING

1. Edge and Fog Computing Integration

Future scheduling frameworks will integrate edge and fog computing to reduce latency and improve decentralized processing.

2. Federated Learning

Federated learning enables distributed ML training without sharing sensitive cloud data directly.

3. AI-Driven Autonomous Clouds

Autonomous cloud infrastructures will use AI agents for self-management and intelligent orchestration.

4. Quantum-Inspired Scheduling

Quantum optimization techniques may improve scheduling efficiency in large-scale distributed systems.

5. Green Cloud Computing

Energy-efficient ML scheduling frameworks will support environmentally sustainable cloud architectures.

Table 3. Literature Review Summary

Author	Technique Used	Major Contribution
Sharma et al.	Reinforcement Learning	Dynamic VM allocation
Kumar and Singh	Deep Learning	Workload prediction model
Verma et al.	Random Forest	Resource utilization optimization
Lee et al.	Hybrid Genetic Algorithm	Multi-objective scheduling
Chen et al.	Q-Learning	Energy-aware cloud scheduling

II. CONCLUSION

Machine learning-based scheduling frameworks represent a transformative advancement in scalable cloud computing architectures. Traditional scheduling methods are insufficient for handling the dynamic and heterogeneous nature of modern cloud environments. Intelligent scheduling systems powered by machine learning improve resource utilization, workload prediction, energy efficiency, scalability, and service quality.

Supervised learning, reinforcement learning, deep learning, and hybrid optimization methods have demonstrated remarkable potential in addressing complex scheduling challenges. Despite several advantages, issues such as computational complexity, security risks, scalability limitations, and real-time adaptation remain significant research challenges.

Future cloud infrastructures are expected to adopt autonomous AI-driven scheduling frameworks integrated with edge computing, federated learning, and green computing technologies. Continued research in intelligent scheduling mechanisms will contribute significantly to the development of highly efficient, scalable, and sustainable cloud computing environments.

REFERENCES

[1]. Sharma, R., & Gupta, P. (2021). Machine learning approaches for cloud task scheduling. *International Journal of Cloud Computing*, 12(3), 145–158.

[2]. Kumar, S., & Singh, A. (2020). Deep learning-based workload prediction in cloud environments. *Journal of Distributed Systems*, 18(2), 89–101.

[3]. Verma, K., (2019). Intelligent resource allocation using random forest algorithms in cloud computing. *International Journal of Computer Applications*, 176(22), 12–20.

[4]. Lee, H., & Park, J. (2022). Hybrid optimization techniques for scalable cloud scheduling systems. *IEEE Transactions on Cloud Computing*, 10(4), 2451–2463.

[5]. Chen, Y., (2021). Reinforcement learning for energy-efficient cloud scheduling. *Future Generation Computer Systems*, 118, 75–88.

[6]. Patel, M., & Shah, D. (2020). Big data analytics and machine learning in cloud scheduling frameworks. *Journal of Big Data Research*, 9(1), 55–70.

[7]. Wang, T., (2019). AI-enabled resource provisioning in distributed cloud architectures. *Cloud Computing Review*, 14(3), 201–214.

[8]. Singh, R., & Kaur, H. (2021). Survey on intelligent scheduling algorithms in cloud computing. *International Journal of Advanced Research in Computer Science*, 11(5), 44–53.