Bulletin of Nexus
Multidisciplinary International Research Journal, Double-Blind Peer Reviewed, Open Access Journal
*ISSN : 3065-7865 / Volume-3, Issue-4, April 2026*

Original Article

# Cloud Task Scheduling Using Metaheuristic Optimization Algorithms: A Review

**V. Sarvani[1], B. Vidyullatha[2]**
[1]Lecturer in Computer Applications, PRR &VS GOVT. College Vidavaluru
[2]Lecturer in Computer Science, PRR &VS GOVT. College Vidavaluru

**Manuscript ID:**
BN-2026-030408

ISSN: 3065-7865

Volume 3

Issue 4

April 2026

Pp. 45-50

**Submitted:** 05 Mar. 2026

**Revised:** 20 Mar. 2026

**Accepted:** 09 Apr. 2026

**Published:** 30 Apr. 2026

DOI:
10.5281/zenodo.21188116
DOI link:
https://doi.org/10.5281/zenodo.21188116

***Abstract***

*Cloud computing is an emerging technology that provides scalable, on-demand computing resources over the internet on a pay-as-you-go basis. It offers a wide range of software and hardware solutions to users, enabling flexible and efficient utilization of computational resources. A critical aspect of cloud computing is task scheduling, which plays a vital role in the virtualization of servers. However, task scheduling is an NP-hard problem, making it computationally challenging to find optimal solutions. To address this, various methodologies have been explored, including heuristic, metaheuristic, and hybrid approaches that combine metaheuristics with machine learning techniques. This paper focuses on the study of key metaheuristic algorithms used in cloud task scheduling, highlighting their strengths, limitations, and the parameters considered during their evaluation. The insights gathered aim to guide future research in developing more effective and adaptive scheduling strategies for cloud environments.*

***Keywords***

*Cloud Computing, Task Scheduling, Multi-Objective Optimization, Metaheuristic Algorithms*

## Introduction

Cloud computing has emerged as a ubiquitous technology that enables users to access a wide range of computing resources over the internet. These resources include servers, storage, applications, and networking components, all delivered as services [1]. At the core of cloud computing lies the concept of virtualization, which abstracts and separates hardware from software, allowing for more flexible and efficient resource management. Task scheduling in cloud computing is an **NP-hard problem**, requiring optimal allocation of tasks based on dynamic network conditions, especially in heterogeneous environments. To address this complexity, various approaches have been proposed, including heuristic**,** metaheuristic**,** and hybrid **methods** that often integrate machine learning techniques [5]**.** Among these, metaheuristic algorithms have gained attention for their ability to provide **near-optimal solutions** efficiently, balancing performance, resource utilization, and system responsiveness in diverse cloud settings [6]. The performance of task scheduling algorithms is measured by their ability to lower makespan, energy consumption and cost, balance workloads more evenly, and improve both throughput and the average utilization of resources [7]. Task scheduling algorithms are broadly classified into heuristics, metaheuristics, and hybrid approaches that integrate machine learning techniques. Metaheuristic algorithms are primarily divided into nature-inspired, physics-based, and swarm intelligence-based methods [8]. This review aims to analyze existing multi-objective scheduling techniques and highlight their strengths, limitations, and future potential.



**Address for correspondence:**
*V. Sarvani, Lecturer in Computer Applications, PRR &VS GOVT. College Vidavaluru*

**How to cite this article:**
*Sarvani, V., & Vidyullatha, B. (2026). Cloud Task Scheduling Using Metaheuristic Optimization Algorithms: A Review. Bulletin of Nexus Journal, 3(4), 45–50. https://doi.org/10.5281/zenodo.21188116*

Quick Response Code:

Website: https://bnir.us

## Background and Fundamentals

### 1 Cloud Computing Architecture

The cloud computing environment primarily involves three components: users who submit task requests, service providers who manage the infrastructure, and data centers where physical resources are housed. Tasks submitted by users are processed and allocated by the system using task scheduling algorithms [3]. Cloud services are generally offered in three main models: **Software as a Service (SaaS)** provides access to software applications and databases via the internet, removing the need for local infrastructure; **Platform as a Service (PaaS)** offers a cloud-based environment for developing, testing, and deploying applications without managing the underlying hardware; and **Infrastructure as a Service (IaaS)** delivers virtualized computing resources such as servers, storage, and networking, allowing users to rent infrastructure on demand[2].

Given the heterogeneous nature of cloud systems—where diverse hardware and software configurations coexist—efficient scheduling algorithms are essential. These algorithms aim to enhance performance, optimize resource utilization, and maintain high-quality service delivery across various operational metrics [4]

### 2 Objectives in Task Scheduling

The main objectives of task scheduling in cloud computing are to minimize the overall completion time (makespan) so that tasks are executed efficiently, reduce the cost of resource usage under the pay-as-you-use model, and improve resource utilization by ensuring that computing resources such as CPU, memory, and virtual machines are used effectively without being idle or overloaded. Additionally, task scheduling aims to minimize energy consumption in data centers to support sustainable and energy-efficient operations, while also ensuring Quality of Service (QoS) by meeting user requirements related to response time, reliability, availability, and throughput in accordance with service-level agreements (SLAs).

## Multi-Objective Optimization Methods: Review

Olanrewaju L. Abraham et al. [1] present a comprehensive Systematic Literature Review (SLR) on multi-objective task scheduling in cloud computing, covering research developments from 2010 to October 2024. The study emphasizes that cloud task scheduling requires efficient allocation of competing tasks to limited resources while simultaneously optimizing multiple objectives such as execution time, cost, energy consumption, and resource utilization. The authors provide a detailed taxonomy and classification of multi-objective optimization techniques, including evolutionary algorithms, swarm intelligence methods, and hybrid approaches, highlighting their growing importance over traditional single-objective methods. Additionally, a comparative analysis of scheduling objectives, simulation environments, evaluation metrics, and datasets used in recent studies is presented, offering valuable insights into current research trends and best-performing approaches. The review concludes that hybrid and metaheuristic techniques are particularly effective in addressing complex scheduling challenges, and it serves as a useful guide for researchers in selecting appropriate methods and tools for designing efficient cloud task scheduling systems.

Mahaboob Basha Shaik et al.[2] propose a novel hybrid optimization approach, PSOSA-LB, for efficient resource allocation in fog-cloud computing environments driven by the rapid growth of Internet of Things (IoT) applications. The algorithm integrates Particle Swarm Optimization (PSO) with Simulated Annealing (SA) to balance exploration and exploitation, where PSO guides the global search while SA helps avoid premature convergence by accepting suboptimal solutions based on a probabilistic mechanism. Additionally, a load balancing component is incorporated through a load imbalance adjustment factor, ensuring even distribution of workloads across fog and cloud resources. The proposed method aims to simultaneously minimize execution time, latency, and energy consumption while improving resource utilization. Experimental results demonstrate that PSOSA-LB significantly outperforms traditional PSO, SA, and other hybrid techniques, achieving up to 33% faster execution, 35% lower latency, 20% reduced energy consumption, and 45% improved load distribution. Overall, the study presents a robust, scalable, and efficient solution for dynamic resource management in fog-cloud environments, making it highly suitable for latency-sensitive and performance-critical IoT applications.

Shanchen Pang et al. [3] address the critical problem of task scheduling in cloud computing, focusing on improving task completion time and system load balancing in large-scale cloud environments. To achieve these objectives, the authors propose a hybrid EDA-GA scheduling algorithm that combines the Estimation of Distribution Algorithm (EDA) with Genetic Algorithm (GA) techniques. The EDA component is utilized to generate feasible solutions through probabilistic modeling and sampling, while GA operators such as crossover and mutation enhance the exploration capability and prevent premature convergence. This integration enables the algorithm to efficiently search for optimal task-to-virtual-machine assignments with improved convergence

speed and solution quality. Experimental evaluation using the CloudSim simulation platform demonstrates that the proposed EDA-GA approach outperforms standalone EDA and GA methods by significantly reducing task completion time and enhancing load balancing performance. Overall, the study presents an effective hybrid optimization strategy that improves scheduling efficiency and resource utilization in cloud computing environments.

Safwat A. Hamad et al. [4] address the critical challenge of resource management in cloud computing, where computational resources such as CPU, memory, storage, and network are provisioned under a virtualization and pay-as-you-go model. The authors propose a Genetic Algorithm (GA)-based task scheduling approach aimed at efficiently allocating application tasks to cloud resources. The main objective of the proposed method is to minimize task completion time and execution cost while maximizing overall resource utilization. By leveraging GA's evolutionary operations such as selection, crossover, and mutation, the algorithm effectively explores the solution space to identify near-optimal scheduling strategies. The performance of the proposed approach is evaluated using the CloudSim simulation toolkit, demonstrating its effectiveness in improving scheduling efficiency compared to traditional methods. Overall, the study highlights the suitability of GA-based optimization techniques for addressing cloud task scheduling problems and enhancing resource utilization in dynamic cloud environments.

Zhenxiang He et al. [5] address the problem of efficient task scheduling in cloud data centers, focusing on improving resource utilization, service quality, and response time. The authors propose a cloudlet scheduling strategy based on the Ant Colony Optimization (ACO) algorithm, where user tasks are treated as the scheduling granularity to optimize virtual machine usage and overall execution time (make-span). The study first explains the fundamental principles and workflow of ACO, and then applies it to design a scheduling strategy that minimizes the number of active virtual machines while maintaining acceptable task execution performance. By leveraging the probabilistic search capability of ACO, the proposed method effectively balances workload distribution and reduces resource wastage in cloud environments. Experimental results conducted using the CloudSim platform demonstrate that the ACO-based approach significantly reduces the number of virtual machines required, lowers operational costs, and achieves a better balance between make-span and service quality when compared to traditional greedy algorithms. Overall, the study highlights the effectiveness of bio-inspired optimization techniques in improving cloud scheduling efficiency.

Arghavan Keivani *et al. [6]* highlight that cloud computing has gained significant attention from large-scale industries due to its flexibility, scalability, and cost-effectiveness. In cloud environments, users typically follow a pay-as-you-go model, where resources are utilized for a specified duration based on demand. This model requires service providers to maintain an optimal balance between resource allocation and pricing in order to ensure profitability while simultaneously satisfying user requirements.

With the rapid growth in the number of user requests, efficient task scheduling has become a challenging issue. The authors emphasize the necessity of robust scheduling frameworks capable of handling large-scale workloads. Scheduling algorithms in cloud computing aim to assign tasks to available resources while maintaining desired Quality of Service (QoS) parameters. Their study presents a comparative analysis of different scheduling techniques, focusing on their effectiveness in optimizing resource utilization and meeting user expectations in dynamic cloud environments.

Fahd Alhaidari *et al [7].* discuss the rapid expansion of cloud computing as an on-demand service model where users are charged based on their resource usage. They emphasize that one of the fundamental challenges in cloud environments is efficient task scheduling, which directly impacts system performance and profitability. The study presents a comparative simulation-based analysis of widely used scheduling algorithms, including First Come First Serve (FCFS), Shortest Task First (STF), Longest Task First (LTF), and Round Robin (RR), using the CloudSim toolkit. The evaluation considers both time-shared and space-shared resource allocation policies. The results indicate that under space-shared allocation, the STF algorithm achieves better performance in terms of total completion time, whereas time-shared allocation demonstrates superior efficiency in reducing overall task execution time compared to space-shared methods. This work highlights the importance of selecting appropriate scheduling strategies based on system requirements and workload characteristics.

Mallu Siva Ramakrishna *et al.* [8] explain that task scheduling in cloud computing involves assigning tasks to virtual machines based on factors such as resource availability, processing capability, memory, and network conditions. They note that although numerous scheduling approaches—

particularly metaheuristic, nature-inspired, and machine learning-based algorithms—have been developed, challenges such as latency persist due to the NP-hard nature of the problem. Their study reviews existing scheduling techniques with respect to key performance parameters, including cost, response time, energy consumption, makespan, throughput, and resource utilization. However, the authors point out that many existing methods overlook critical aspects such as trust and fault tolerance. They emphasize that trust plays an essential role in reliable task execution, while fault tolerance ensures system continuity in the presence of failures. The study concludes that incorporating both trust and fault tolerance can significantly enhance Quality of Service (QoS) and improve overall scheduling efficiency, and it suggests future research directions focusing on these aspects.

Arghavan Keivani *et al.* [9] observe that cloud computing has attracted significant interest from large enterprises due to its scalability, flexibility, and economic benefits. In this model, users are charged based on resource usage over a specified time period, making cost management a critical factor. The authors emphasize that service providers must maintain an effective balance between resource allocation and pricing to ensure profitability while meeting user expectations. With the rapid increase in user demands, efficient task scheduling becomes a major challenge, necessitating the development of robust scheduling frameworks. Scheduling algorithms are designed to allocate tasks to available resources while considering Quality of Service (QoS) requirements. The study provides a comparative review of various cloud task scheduling techniques, highlighting their effectiveness in improving resource utilization and overall system performance.

Farida *et al.* [10] describe that the emergence of cloud computing has enabled organizations to efficiently deploy and manage applications with easy access to a wide range of services. They emphasize the importance of effective and dynamic task scheduling to handle multiple user requests in a heterogeneous resource environment. Inefficient scheduling may result in resource underutilization or overutilization, ultimately affecting system performance. The authors highlight the effectiveness of nature-inspired optimization techniques in addressing scheduling challenges in cloud computing. Their study presents a comprehensive review and classification of such techniques based on factors including scheduling algorithms, problem characteristics, task types, optimization objectives, resource mapping strategies, constraints, and evaluation environments. Additionally, the work outlines future research directions, providing valuable insights for researchers and practitioners in advancing cloud task scheduling methodologies.

Yellamma Pachipala *et al.* [11] emphasize the importance of efficient task scheduling in cloud computing for improving resource utilization and overall system performance. They propose a modified version of the Shortest Job First (SJF) scheduling algorithm implemented using the CloudSim simulation framework. The study aims to overcome limitations of traditional scheduling techniques by reducing resource bottlenecks and minimizing task completion time. Their results indicate that the proposed modified SJF approach enhances system efficiency and provides better performance in dynamic cloud environments.

Boonhatai Kruekaew *et al.[12]* address the challenge of workload balancing in cloud computing, particularly within Infrastructure as a Service (IaaS) environments, where improper load distribution can lead to server overloading or underutilization, affecting system performance and reliability. To overcome these issues, the authors propose a multi-objective task scheduling approach that integrates the Artificial Bee Colony (ABC) algorithm with Q-learning, referred to as the MOABCQ method. This hybrid technique aims to optimize multiple objectives, including makespan, cost, resource utilization, and virtual machine (VM) throughput, while ensuring effective load balancing across VMs. The performance of the proposed method is evaluated using CloudSim and compared with existing algorithms such as Max-Min, FCFS, HABC_LJF, Q-learning, MOPSO, and MOCS across various datasets, including Random, Google Cloud Jobs, and synthetic workloads. The results demonstrate that the MOABCQ approach significantly outperforms traditional methods in reducing makespan and cost, minimizing load imbalance, and improving throughput and overall resource utilization.

Shubham Mittal *et al.[13]* describe cloud computing as an on-demand service model that enables users to access computing resources such as storage, servers, networks, and applications remotely with ease. They emphasize that efficient execution of tasks in cloud environments requires effective scheduling mechanisms to ensure optimal resource utilization, reduced makespan, and improved system performance. To address these challenges, the authors propose an optimized task scheduling algorithm that dynamically adapts features from existing algorithms based on system conditions. The proposed approach considers the distributed and scalable nature of cloud resources, thereby

enhancing task allocation efficiency and overall performance in cloud computing environments.

R. K. Jena [14] highlights that cloud computing consists of a collection of heterogeneous and autonomous systems with flexible computational capabilities. The author emphasizes that task scheduling is a critical component for enhancing overall system performance, reducing processing time, and minimizing energy consumption, which in turn improves the profitability of service providers. To address these objectives, the study proposes a multi-objective nested Particle Swarm Optimization (TSPSO) algorithm that simultaneously optimizes energy usage and execution time. The proposed approach is implemented and evaluated using the CloudSim simulation platform. Experimental results demonstrate that the TSPSO algorithm achieves a better balance among multiple objectives compared to existing scheduling techniques, thereby improving efficiency in cloud computing environments.

**Conclusion**

Multi-objective optimization plays a crucial role in enhancing cloud task scheduling by effectively balancing conflicting factors such as execution time, cost, and energy consumption. Numerous algorithms have been developed, each offering distinct advantages while also presenting certain limitations. In recent years, hybrid techniques and artificial intelligence-based approaches have shown considerable potential in addressing complex scheduling challenges. However, further research is required to overcome issues related to scalability, dynamic workload management, and energy efficiency in evolving cloud environments.

**Acknowledgment**

The author(s) express their sincere gratitude to all those who contributed to the successful completion of this review paper entitled **"Cloud Task Scheduling Using Metaheuristic Optimization Algorithms: A Review."**

We are deeply thankful to our guide, faculty members, and academic mentors for their valuable guidance, constructive suggestions, and continuous encouragement throughout the preparation of this work. Their expertise and support greatly enhanced the quality of the study.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

The authors declare that there are no conflicts of interest regarding the publication of this paper.

**References**

1. O. L. Abraham, M. A. B. Ngadi, J. B. M. Sharif, and M. K. M. Sidik, "Multi-objective optimization techniques in cloud task scheduling: A systematic literature review," *IEEE Access*, 2025.
2. M. B. Shaik, K. S. Reddy, K. Chokkanathan, S. A. A. Biabani, P. Shanmugaraja, and D. R. D. Brabin, "A hybrid particle swarm optimization and simulated annealing with load balancing mechanism for resource allocation in fog-cloud environments," *IEEE Access*, early access, 2024, doi: 10.1109/ACCESS.2024.3489960.
3. S. Pang, W. Li, H. He, Z. Shan, and X. Wang, "An EDA-GA hybrid algorithm for multi-objective task scheduling in cloud computing," *IEEE Access*, vol. 7, pp. 134517–134528, 2019, doi: 10.1109/ACCESS.2019.2946216.
4. S. A. Hamad and F. A. Omara, "Genetic-based task scheduling algorithm in cloud computing environment," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 7, no. 4, 2016.
5. Z. He, J. Dong, Z. Li, and W. Guo, "Research on task scheduling strategy optimization based on ACO in cloud computing environment," in *Proc. IEEE 5th Int. Conf. Inf. Technol. Mechatronics Eng. (ITOEC)*, 2020.
6. A. Keivani, F. Ghayoor, and J.-R. Tapamo, "A review of recent methods of task scheduling in cloud computing," in *Proc. IEEE Int. Conf.*, 2018.
7. F. Alhaidari, T. Balharith, and E. Al-Yahyan, "Comparative analysis for task scheduling algorithms on cloud computing," Imam Abdulrahman Bin Faisal University, Saudi Arabia, 2018.
8. M. S. R. Krishna and S. Mangalampalli, "A systematic review on various task scheduling algorithms in cloud computing," *EAI Endorsed Trans. Internet of Things*, Review Article.
9. F. S. Prity, K. M. A. Uddin, et al., "A review of task scheduling in cloud computing based on nature-inspired optimization algorithm," *Cluster Computing*, 2023, doi: 10.1007/s10586-023-04090-y.
10. Y. Pachipala et al., "Optimizing task scheduling in cloud computing: An enhanced shortest job first algorithm," in *Proc. Int. Conf. Intell. Data Commun. Appl. (ICIDCA)*, Elsevier, 2024.
11. B. Kruekaew and W. Kimpan, "Multi-objective task scheduling optimization for load balancing in cloud computing environment using hybrid artificial bee colony algorithm with reinforcement learning," *IEEE Access*, vol. 10, pp. 13724–13738, 2022, doi: 10.1109/ACCESS.2022.3149955.

12. S. Mittal and A. Katal, "An optimized task scheduling algorithm in cloud computing," in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, 2016, doi: 10.1109/IACC.2016.45.
13. R. K. Jena, "Multi-objective task scheduling in cloud environment using nested PSO framework," in *Proc. 3rd Int. Conf. Recent Trends Comput. (ICRTC)*, Elsevier, 2015.
14. X. Yu et al., "Dynamic multi-objective task scheduling in cloud computing using reinforcement learning for energy and cost optimization," *Scientific Reports*, 2025.
15. O. L. Abraham et al., "Multi-objective optimization techniques in cloud task scheduling: A systematic literature review," *IEEE Access*, 2025.
16. V. K. Vasantham and H. Donavalli, "Multi-objective hybrid optimized task scheduling in cloud computing under big data perspective," 2024.
17. Q. Li et al., "UDL: A cloud task scheduling framework based on multiple deep neural networks," *Journal of Cloud Computing*, 2023.
18. R. Bhakhar and R. S. Chhillar, "Dynamic multi-criteria scheduling algorithm for smart home tasks in fog-cloud IoT systems," *Scientific Reports*, 2024.
19. Z. R. M. Azmi et al., "Comparative analysis of task scheduling in multi-tier fog-cloud computing," 2026.
20. L. Yan, "QoS-aware task scheduling using modified wombat optimization algorithm," 2025.
21. A. Selvaraj et al., "Multi-objective job scheduling in green cloud with big data," 2025.