

A Survey of Cloud Computing Resource Scheduling Algorithms

RABINA BAGGA • KAMALI GUPTA

Abstract. A structural design that involves virtual machines connecting to a cloud service provider is called cloud computing. Cloud Computing is the most recent new trend to put infrastructure resources' visions into action. Cloud computing is a technological innovation that builds on how computing systems develop technologies besides using established software-building resources. It is based on the flexible provisioning principle; this term encompasses services and machines, storage, networking, and information technology (IT) networks in general. Cloud Computing has developed into a model of commoditized services.

Keywords: Cloud Computing, Load Balancing, Optimization Algorithms, Fuzzy Logic.

INTRODUCTION

The widespread adoption of cloud computing across various sectors stems from the increasing ubiquity of web connectivity and the exponential growth of big data in terms of its volume, speed, and diversity online. This trend introduces a fresh paradigm in computing resources through the cloud, offering the range of services such as IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service) to users worldwide, regardless of their location or device [Wei19]. These resources are accessible online, providing flexibility through on-demand access and payment models such as pay-per-use or subscriptions. Cloud computing, essentially a federation of numerous resources, employs virtualization for resource management, akin to parallel computing, yet distinguished by its composition of diverse resources. This enables the provision of computing resources to users as a utility, often referred to as the fifth utility, akin to traditional services like electricity, gas, water, and telephone service for instance, highly scalable computing nodes are now easily accessible, anytime and anywhere, and users only incur charges for their actual usage. Such flexibility and scalability would have been unattainable with conventional data centers or traditional computing resources [Pha17]. Virtualization blurs the conventional boundaries of computing resources, ushering in a new realm of research that is both influenced and propelled by the increasing demand for cloud computing. Given its utility-oriented nature, cloud computing allows providers and customers to concentrate on their core businesses, optimizing profits and return on investment through strategic resource scheduling. Effective resource allocation is particularly critical for cloud computing services, as it directly impacts accessibility, anytime and anywhere, and users only incur charges for their actual usage. Such flexibility and scalability would have been unattainable with conventional data centers or traditional computing resources [Pha17].

Virtualization blurs the conventional boundaries of computing resources, ushering in a new realm of research that is both influenced and propelled by the increasing demand for cloud computing. Given its utility-oriented nature, cloud computing allows providers and customers to concentrate on their core businesses, optimizing profits and return on investment through strategic resource scheduling. Effective resource allocation is particularly critical for cloud computing services, as it directly impacts the ability to meet SLA (Service Level Agreement). Underestimating available resources would result in SLA breaches and fines [Bit18], while overestimating them might result in resource underutilization and revenue loss. Cloud resource scheduling presents challenges for providers aiming for efficiency in resource and power usage. Moreover, it poses a significant dilemma for users concerning the QoS (Quality of Service) standards within the pay-as-you-go provision.

Cloud Computing

Each cloud computing application was typically structured as a business process comprising multiple abstract services. These abstract services encapsulate the functionalities of various application components through their interfaces, while concrete services or resources are dynamically allocated at runtime to fulfill these functions. Conceptually, users acquire IT infrastructures or computing platforms from cloud environments, where they deploy and execute their applications. Therefore, cloud computing provides users with uninterrupted access to hardware, software and information services and provides a unified computing platform [Pat17]. Cloud computing generally provides mainly three types of services:

1) *IaaS (Infrastructure as a Service)*: Due to rapid advances in infrastructure virtualization, IT automation, usage metering and pricing, users can now procure IT infrastructure or even entire data centers on a pay-per-use subscription model. To meet user expectations, IaaS (Infrastructure as a Service) offerings are designed to be flexible, scalable, and managed efficiently [Aru19].

2) *PaaS (Platform as a Service)*: Services encompass both the solution stack and the computing platform. PaaS (Platform as a Service) represents a category of cloud computing positioned between SaaS (Software as a Service) and IaaS (Infrastructure as a Service). PaaS enables users to deliver complete operating systems and associated services over the internet without requiring downloads or installations. Often termed "cloudware," PaaS offers the flexibility to migrate from network cloud resources to personal PCs. It builds upon the foundation of Software as a Service and provides clients with access to hosted software applications over the internet

3) *SaaS (Software as a Service)*: Customers can access software or applications as a hosted service over the Internet, eliminating the need to install and run applications on personal computers [Alk16]. Through this approach, customers benefit from on-demand pricing, alleviating the burden of software maintenance and reducing software procurement costs. One of the earliest manifestations of SaaS (Software as a Service) is the ASP (Application Service Provider) model, where software hosted or distributed online can be subscribed to via the ASP method.

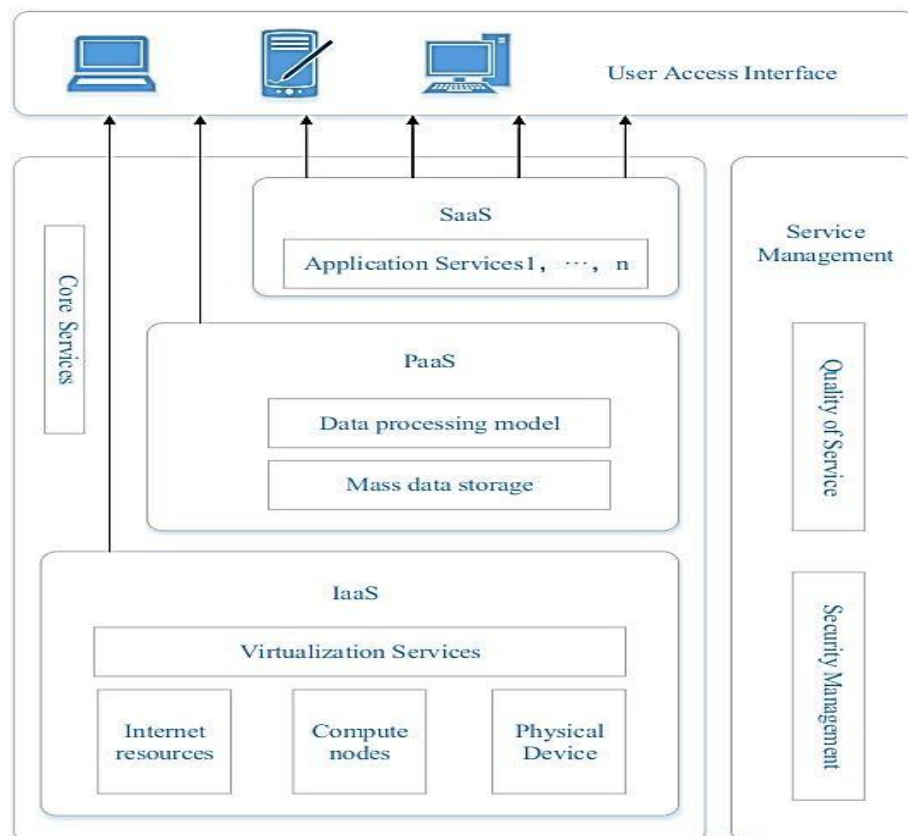


Fig. 1 Cloud computing architecture

Cloud computing provides a set of on-demand services that enable scalable resource allocation. Its design can be divided into three main layers: core services, service management, and user interface, which align with current research and applications in the field [Hu18]. The Core Services layer includes hardware infrastructure, software operating environments, and application abstraction, which provide scalability, high availability, and reliability to accommodate various application needs. Service management focuses on improving the dependability, accessibility, and security of these critical services. End users interact with the cloud through the user interface layer [Aga18].

Cloud Resource Scheduling Problem

Cloud computing boasts distinctive traits such as virtualization, heterogeneity, metered service and pricing, elasticity, and resource pooling, despite its roots in established paradigms like cluster and grid computing. However, achieving these unique features presents various challenges, including security and privacy concerns, resource scheduling, scalability and fault tolerance, energy conservation, and interoperability. Resource providers must address [Ben18] these challenges while delivering services to customers to maintain consumer trust in the cloud. While consumers seek cost-effective, efficient, and reliable services, service providers goal to optimize their profit and return on investment.

In order to meet the objectives of both consumers and providers, a robust mechanism for resource management is essential. When referring to resources, we encompass any component capable of serving a user, including logical or physical entities such as operating systems, energy, bandwidth, and throughput, as well as physical resources like processors, memory, storage, and network components. Effective resource management involves scheduling and provisioning these resources efficiently [Bu19]. Resource provisioning involves identifying, selecting, and allocating the necessary resources to execute activities and workflows appropriately. On the other hand, resource scheduling entails creating a schedule that determines the assignment of tasks to specific resources. Both resource provisioning and scheduling are essential for ensuring SLAs (Service Level Agreements) are upheld and QoS (Quality of Service) requirements are met [Ran19]. Fig 2 illustrates the process of resource scheduling in cloud computing.

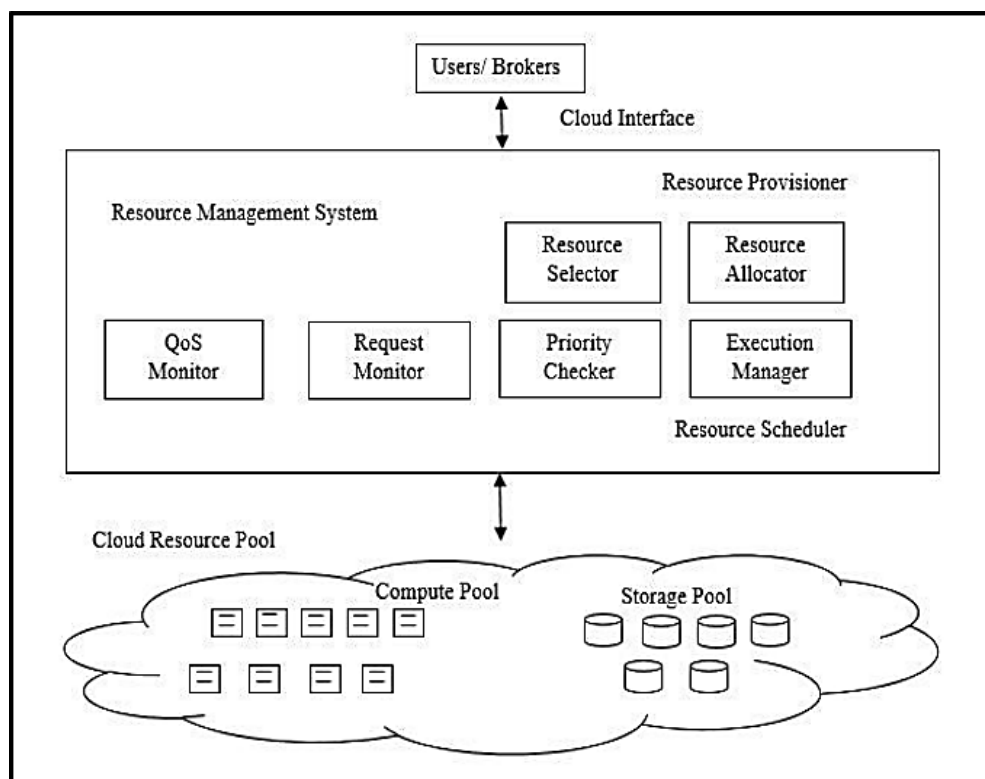


Fig 2 Cloud Computing Resource Scheduling Problem

Tasks and workflows are sent to the cloud computing environment by the users or an agent on behalf of the user via cloud interface. The RMS (Resource Management System) is responsible for monitoring the status of submitted tasks [Lu20], determining resource requirements, ensuring SLA compliance, and overseeing task completion. Once resources are allocated by the resource provisioner, the resource scheduler is invoked by the RMS for task execution. The resource scheduler comprises several components, including the execution manager, priority checker, request monitor, and QoS monitor, which interact to manage resource allocation efficiently [Gaz20].

Scheduling Virtual Resources in Cloud

Clouds can also be considered as virtual resource pools comprising hardware, development platforms, and services that are easily accessible and utilized. Currently, cloud computing stands out as the most promising large-scale computing solution, primarily due to the cost-effectiveness and user-friendliness enabled by resource virtualization technology. While cloud computing has greatly benefited from virtualization, its scheduling mechanisms differ from those of traditional distributed systems [Yad20]. In traditional distributed systems, as depicted in Fig. 3, tasks are sent to physical devices by the scheduler upon reception, following a one-to-one scheduling approach. Conversely, Fig. 4 illustrates the scheduling decisions in cloud computing systems.

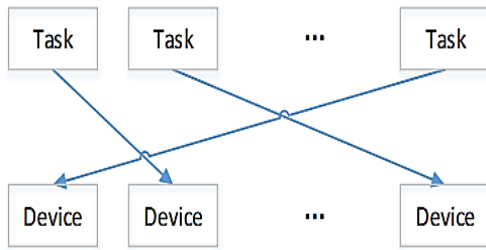


Fig. 3 Distributed resource scheduling model

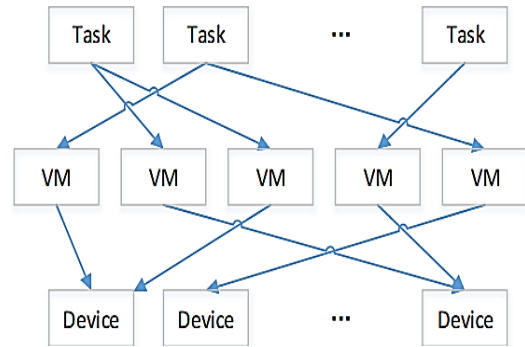


Fig. 4 Cloud computing resource scheduling model

After receiving a request from users, the system should promptly allocate virtual resources to respond effectively. When a job requires multiple virtual machines, these machines are shared on a single physical device, and the scheduler assigns the appropriate virtual machines to the device. The introduction of virtualization technology has thus facilitated more efficient resource utilization in cloud computing system resource scheduling [Bit18b].

The RSP (Resource Scheduling Problem) can be viewed as a function that maps jobs or tasks (represented by a set $x_1, x_2, x_3, \dots, x_n$) to resources (represented by a set of VMs (Virtual Machines) $v_1, v_2, v_3, \dots, v_m$) in a manner that optimizes the objective function, whether it's minimizing or maximizing it. These tasks can either be independent or interconnected. The objective function's behavior is determined by SLA agreements and user requirements. The objective function may depend on many parameters, ranging from single to multiple ones. Each parameter might contribute to the objective function individually or through distinct weights assigned to them [Pen20].

The RSP is known as an NP-hard problem, indicating that as the dimensions and complexity increase, these problems cannot be solved in polynomial time. To find improved solutions with faster convergence rates, researchers have turned to evolutionary techniques such as genetic algorithms, particle swarm optimization, and ant colony optimization. These methods offer efficient strategies for exploring the solution space and finding near-optimal solutions within a reasonable time frame [Che20].

Literature Review

Author (Year)	Technique Used	Simulation Tool	Parameters	Results	Advantages	Limitations
P. V. Reddy, et.al (2023)	Multi-objective scheduling (MOS) mechanism	Cloudsim	Makespan, security and resource availability	The simulation results depicted that the suggested mechanism was more effective with third algorithm in comparison with others.	This mechanism offered higher security and lower makespan.	This mechanism was not scheduled workflow in manifold clouds using real-time data.
S. Singhal, et.al (2023)	Rock Hyrax-based optimization method	CloudSim 3.0.3	Makespan, energy efficiency, response time, throughput, and cost	Based on findings, the introduced method was useful for enhancing QoS metrics up to 3%-8% in comparison with other methods concerning algorithms.	This method had worked robustly in the case of dynamic jobs and resources.	This method was inapplicable in a real cloud environment and consumed higher transmitting jobs and data.
S. Prathiba, et.al (2023)	L3F-MGA and E-ANFIS algorithms	JAVA platform	Processing time and energy consumption (EC)	The experimental results exhibited the supremacy of designed algorithms over traditional methods.	The initial algorithm led to mitigating energy consumption (EC) up to 11% and the latter one alleviated the processing time of up to 8% in contrast to CM-GA.	These algorithms were not able to adjust data center configurations in a dynamic way considering energy availability and environmental circumstances.
A. Moazeni, et.al (2023)	Adaptive multi-objective teaching-learning based optimization (AMO-TLBO) algorithm	CloudSim	Makespan, and cost	The simulations demonstrated that the presented algorithm was more effective in contrast to traditional methods concerning diverse parameters.	The presented algorithm had alleviated the makespan, cost and enhanced its deployment with well-balanced load over virtual machines (VM).	This algorithm consumed higher energy.
M. Radi, et.al (2023)	Modified genetic-based VM consolidation (MGVMC) method	CloudSim Plus	Delay and makespan	The simulations demonstrated that the presented method was robust to mitigate energy consumption, SLA violations, and the amount of VM migrations than conventional methods.	This method was effective to optimize VM consolidation in the cloud environment.	Some issues related to process cloud data were not resolved.
P. V. Lahande, et.al (2023)	Load Balancing (LB) technique	Workflow-Sim	Resource usage, delay, makespan and throughput	The developed technique balanced the load up to 51.98% with the first algorithm, 41.71% with second, 51.98% with third, 59.43% with fourth, and 52.17% with last algorithm.	The Reinforcement Learning (RL) technique was useful for balancing the load and optimizing the cloud resource usage, and offered higher quality of service (QoS), such as delay, makespan and throughput.	This technique was ineffective in real-time applications.

Author (Year)	Technique Used	Simulation Tool	Parameters	Results	Advantages	Limitations
G. Senthilku- mar, et.al (2023)	A hybrid method called GA- RF	CloudSim 3.0	Resource usage, energy utilization, and execution time	The experimental results exhibited that the projected method was performed well against existing technique and offered lower resource usage, energy consumption and execution time.	The supremacy of initial algorithm was proved over others.	This method became less reliable and less scalable in real- time cloud settings.
M. S. Al Reshan, et.al (2023)	GWO-PSO approach	MATLAB	Response time and convergence	The experiments validated that the presented approach was effective for alleviating entire response time up to 12% and offered convergence of 97.253%.	The presented approach was feasible to optimize the fast convergence and mitigate the entire response time as compared to other methods.	This approach led to avoiding the perfect balance of examination and utilization due to which the performance of both models was lessened.
N. K. Sharma, et.al (2023)	BPEVA and MDVMM	Cloudsim	Energy consumption, resource usage	The recommended algorithms were effectual to save energy usage up to 31% and enhance 21.7% average resource usage at confidence interval (CI) of 95%.	These algorithms were performed well concerning SLA violation, VMs migration, and Energy SLA Violation (ESV).	These algorithms were ineffective for predicting over- and under-usage state of PMs.
W. Gu, et.al (2024)	Mobile Edge Caching based Resource Scheduling (MECRS) technique	MATLAB	Service rate	The simulations depicted the effectiveness of the suggested technique regarding higher service rates and restricted number of resources.	This technique ensured that other requests were suitable and offered a service rate of 50%.	Other issues related to detecting vehicle faults were not resolved using this technique.

CONCLUSION

Load balancing algorithms can be divided into two main types: dynamic and static. In static algorithms, decisions about load shifting are made without considering the current stage of the system; these algorithms require awareness of the resources and applications in the system. Once tasks are assigned, the performance of the VMs (virtual machines) can be measured. The master node then distributes tasks to slave nodes based on their performance, with the slave processors handling the assigned work and the master process provides the results. In contrast, dynamic algorithms make load balancing decisions based on the current state of the system. The dynamic algorithms give high performance as compared to static algorithms for the load balancing. In future a scheme of dynamic load balancing will be proposed for the load balancing in cloud computing.

REFERENCES

- [Aga18] M. Agarwal and G. M. S. Srivastava, "Genetic algorithm-enabled particle swarm optimization (PSOGA)-based task scheduling in cloud computing environment," *Int. J. Inf. Technol. Decis. Making*, vol. 17, no. 4, pp. 1237-1267, Jul. 2018. DOI: [10.1142/S0219622018500244](https://doi.org/10.1142/S0219622018500244).
- [Alk16] E. S. Alkayal, N. R. Jennings, and M. F. Abulkhair, "Efficient task scheduling multi-objective particle swarm optimization in cloud computing," in *Proc. IEEE 41st Conf. Local Comput. Netw. Workshops (LCN Workshops)*, Nov. 2016, pp. 17-24. DOI: [10.1109/LCN.2016.024](https://doi.org/10.1109/LCN.2016.024).
- [Aru19] A. Arunarani, D. Manjula, and V. Sugumaran, "Task scheduling techniques in cloud computing: A literature survey," *Future Gener. Comput. Syst.*, vol. 91, pp. 407-415, Feb. 2019. DOI: [10.1016/j.future.2018.09.014](https://doi.org/10.1016/j.future.2018.09.014).

- [Ben18] H. Ben Alla, S. Ben Alla, A. Touhafi, and A. Ezzati, "A novel task scheduling approach based on dynamic queues and hybrid meta-heuristic algorithms for cloud computing environment," *Cluster Comput.*, vol. 21, no. 4, pp. 1797-1820, Dec. 2018. DOI: [10.1007/s10586-018-2811-x](https://doi.org/10.1007/s10586-018-2811-x). EDN: [WDQDYD](https://www.edn.net/WDQDYD).
- [Bit18] L. F. Bittencourt, A. Goldman, E. R. M. Madeira, N. L. S. da Fonseca, and R. Sakellariou, "Scheduling in distributed systems: A cloud computing perspective," *Comput. Sci. Rev.*, vol. 30, pp. 31-54, Nov. 2018. DOI: [10.1016/j.cosrev.2018.08.002](https://doi.org/10.1016/j.cosrev.2018.08.002). EDN: [CEHYCM](https://www.edn.net/CEHYCM).
- [Bit18b] C. Bitsakos, I. Konstantinou, and N. Koziris, "DERP: A deep reinforcement learning cloud system for elastic resource provisioning," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci. (CloudCom)*, Dec. 2018, pp. 21-29. DOI: [10.1109/CloudCom2018.2018.00020](https://doi.org/10.1109/CloudCom2018.2018.00020).
- [Bu19] F. Bu and X. Wang, "A smart agriculture IoT system based on deep reinforcement learning," *Future Gener. Comput. Syst.*, vol. 99, pp. 500-507, Oct. 2019. DOI: [10.1016/j.future.2019.04.041](https://doi.org/10.1016/j.future.2019.04.041).
- [Car19] F. Carpio, A. Jukan, R. Sosa, and A. J. Ferrer, "Engineering a QoS provider mechanism for edge computing with deep reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1-6. DOI: [10.1109/GLOBECOM38437.2019.9013946](https://doi.org/10.1109/GLOBECOM38437.2019.9013946).
- [Che18] M. Cheng, J. Li, and S. Nazarian, "DRL-cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers," in *Proc. 23rd Asia South Pacific Design Autom. Conf. (ASPDAC)*, Jeju, South Korea, Jan. 2018, pp. 129-134. DOI: [10.1109/ASPDAC.2018.8297294](https://doi.org/10.1109/ASPDAC.2018.8297294).
- [Che20] M. Cheng, J. Li, P. Bogdan, and S. Nazarian, "H2O-cloud: A resource and quality of service-aware task scheduling framework for warehouse-scale data centers," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2925-2937, Oct. 2020. DOI: [10.1109/TCAD.2019.2930575](https://doi.org/10.1109/TCAD.2019.2930575). EDN: [VGZCUB](https://www.edn.net/VGZCUB).
- [Che20b] M. Cheng, J. Li, P. Bogdan, and S. Nazarian, "H2O-cloud: A resource and quality of service-aware task scheduling framework for warehouse-scale data centers," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2925-2937, Oct. 2020. DOI: [10.1109/TCAD.2019.2930575](https://doi.org/10.1109/TCAD.2019.2930575). EDN: [VGZCUB](https://www.edn.net/VGZCUB).
- [Gaz20] P. Gazori, D. Rahbari, and M. Nickray, "Saving time and cost on the scheduling of fog-based IoT applications using deep reinforcement learning approach," *Future Gener. Comput. Syst.*, vol. 110, pp. 1098-1115, Sep. 2020. DOI: [10.1016/j.future.2019.09.060](https://doi.org/10.1016/j.future.2019.09.060). EDN: [TYAXQA](https://www.edn.net/TYAXQA).
- [Hu18] H. Hu, Z. Li, H. Hu, J. Chen, J. Ge, C. Li, and V. Chang, "Multi-objective scheduling for scientific workflow in multicloud environment," *J. Netw. Comput. Appl.*, vol. 114, pp. 108-122, Jul. 2018. DOI: [10.1016/j.jnca.2018.03.028](https://doi.org/10.1016/j.jnca.2018.03.028).
- [Li17] H. Li, J. Li, W. Yao, S. Nazarian, X. Lin, and Y. Wang, "Fast and energy-aware resource provisioning and task scheduling for cloud systems," in *Proc. 18th Int. Symp. Qua.* DOI: [10.1109/ISQED.2017.7918312](https://doi.org/10.1109/ISQED.2017.7918312).
- [Li18] H. Li, R. Cai, N. Liu, X. Lin, and Y. Wang, "Deep reinforcement learning: Algorithm, applications, and ultra-low-power implementation," *Nano Commun. Netw.*, vol. 16, pp. 81-90, Jun. 2018. DOI: [10.1016/j.nancom.2018.02.003](https://doi.org/10.1016/j.nancom.2018.02.003).
- [Liu17] N. Liu, Z. Li, J. Xu, Z. Xu, S. Lin, Q. Qiu, J. Tang, and Y. Wang, "A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Atlanta, GA, USA, Jun. 2017, pp. 372-382. DOI: [10.1109/ICDCS.2017.123](https://doi.org/10.1109/ICDCS.2017.123).
- [Lu20] H. Lu, C. Gu, F. Luo, W. Ding, and X. Liu, "Optimization of lightweight task offloading strategy for mobile edge computing based on deep reinforcement learning," *Future Gener. Comput. Syst.*, vol. 102, pp. 847-861, Jan. 2020. DOI: [10.1016/j.future.2019.07.019](https://doi.org/10.1016/j.future.2019.07.019). EDN: [EZINJI](https://www.edn.net/EZINJI).
- [Nou19] S. M. R. Nouri, H. Li, S. Venugopal, W. Guo, M. He, and W. Tian, "Autonomic decentralized elasticity based on a reinforcement learning controller for cloud applications," *Future Gener. Comput. Syst.*, vol. 94, pp. 765-780, May 2019. DOI: [10.1016/j.future.2018.11.049](https://doi.org/10.1016/j.future.2018.11.049).
- [Pat17] N. Patil and D. Aeloor, "A review-different scheduling algorithms in cloud computing environment," in *Proc. 11th Int. Conf. Intell. Syst. Control (ISCO)*, Coimbatore, India, Jan. 2017, pp. 182-185. DOI: [10.1109/ISCO.2017.7855977](https://doi.org/10.1109/ISCO.2017.7855977).
- [Pen20] Z. Peng, J. Lin, D. Cui, Q. Li, and J. He, "A multi-objective trade-off framework for cloud resource scheduling based on the deep Q-network algorithm," *Cluster Comput.*, Jan. 2020, DOI: [10.1007/s10586-019-03042-9](https://doi.org/10.1007/s10586-019-03042-9).
- [Pha17] S. Phaniteja, P. Dewangan, P. Guhan, A. Sarkar, and K. M. Krishna, "A deep reinforcement learning approach for dynamically stable inverse kinematics of humanoid robots," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Macau, China, Dec. 2017, pp. 1818-1823. DOI: [10.1109/ROBIO.2017.8324682](https://doi.org/10.1109/ROBIO.2017.8324682). EDN: [VIKSMI](https://www.edn.net/VIKSMI).
- [Qua18] L. Quan, Z. Wang, and F. Ren, "A novel two-layered reinforcement learning for task offloading with tradeoff between physical machine utilization rate and delay," *Future Internet*, vol. 10, no. 7, pp. 10-60, 2018. DOI: [10.3390/fi10070060](https://doi.org/10.3390/fi10070060).
- [Ran19] L. Ran, X. Shi, and M. Shang, "SLAs-aware online task scheduling based on deep reinforcement learning method in cloud environment," in *Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun.*, Zhangjiajie, China, Aug. 2019, pp. 1518-1525. DOI: [10.1109/HPCC/SmartCity/DSS.2019.00209](https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00209).

- [Wan19] J. Wang, L. Zhao, J. Liu, and N. Kato, "Smart resource allocation for mobile edge computing: A deep reinforcement learning approach," *IEEE Trans. Emerg. Topics Comput.*, early access, Mar. 4, 2019, DOI: [10.1109/TETC.2019.2902661](https://doi.org/10.1109/TETC.2019.2902661). EDN: WYSGAE.
- [Wei19] Y. Wei, D. Kudenko, S. Liu, L. Pan, L. Wu, and X. Meng, "A reinforcement learning based auto-scaling approach for SaaS providers in dynamic cloud environment," *Math. Problems Eng.*, vol. 2019, pp. 1-11, Feb. 2019. DOI: [10.1155/2019/5080647](https://doi.org/10.1155/2019/5080647). EDN: QJMPEU.
- [Xia19] A. Xiao, Z. Lu, J. Li, J. Wu, and P. C. K. Hung, "SARA: Stably and quickly find optimal cloud configurations for heterogeneous big data workloads," *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105759. DOI: [10.1016/j.asoc.2019.105759](https://doi.org/10.1016/j.asoc.2019.105759).
- [Yad17] R. Yadav and W. Zhang, "MeReg: Managing energy-SLA tradeoff for green mobile cloud computing," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1-11, Dec. 2017. DOI: [10.1155/2017/6741972](https://doi.org/10.1155/2017/6741972).
- [Yad18] R. Yadav, W. Zhang, et al., "Adaptive energy-aware algorithms for minimizing energy consumption and SLA violation in cloud computing," *IEEE Access*, vol. 6, pp. 55923-55936, 2018. DOI: [10.1109/ACCESS.2018.287275](https://doi.org/10.1109/ACCESS.2018.287275).
- [Yad20] R. Yadav, W. Zhang, K. Li, C. Liu, M. Shafiq, and N. K. Karn, "An adaptive heuristic for managing energy consumption and overloaded hosts in a cloud data center," *Wireless Netw.*, vol. 26, no. 3, pp. 1905-1919, Apr. 2020. DOI: [10.1007/s11276-018-1874-1](https://doi.org/10.1007/s11276-018-1874-1). EDN: BOOTKQ.

МЕТАДАТА | МЕТАДАННЫЕ

The article was received by the editors on January 27, 2025

Поступила в редакцию 27 января 2025 г.

Название: Обзор алгоритмов планирования ресурсов облачных вычислений.

Аннотация: Структурная конструкция, включающая виртуальные машины, подключающиеся к поставщику облачных услуг, называется облачными вычислениями. Облачные вычисления — это новейшая тенденция, позволяющая реализовать видение ресурсов инфраструктуры. Облачные вычисления — это технологическое новшество, которое основывается на том, как вычислительные системы разрабатывают технологии, помимо использования существующих ресурсов для создания программного обеспечения. Оно основано на принципе гибкого предоставления; этот термин охватывает службы и машины, хранилища, сети и сети информационных технологий (ИТ) в целом. Облачные вычисления превратились в модель товаров и услуг.

Ключевые слова: Облачные вычисления, балансировка нагрузки, алгоритмы оптимизации, нечеткая логика.

Язык статьи: Английский.

About the authors | Об авторах

Rabina Bagga

Chitkara University, India.
Research Scholar (Ph.D), in the Department of Computer Science and Engineering. Her research interests include cloud computing, machine learning algorithms, and data structures.
E-mail: rabinabagga1005cse.phd22@chitkara.edu.in

Kamali Gupta

Chitkara University, India.
Professor, DCSE. She has guided many M.Tech thesis and has undergone numerous MOOC course certifications. Her research interests include cloud computing, security, computer architecture and data structures.
E-mail: kamali.gupta@chitkara.edu.in

Рабина Багга

Университет Читкара, Индия.
Научный сотрудник (Ph.D) на кафедре компьютерных наук и инженерии. Иссл. интересы: облачные вычисления, алгоритмы машинного обучения и структуры данных.
E-mail: rabinabagga1005cse.phd22@chitkara.edu.in

Камали Гупта

Университет Читкара, Индия.
Профессор, DCSE. Руководила магистерскими диссертациями, сертифицирована по онлайн-курсам. Иссл. интересы: облачные вычисления, безопасность, компьютерная архитектура и структуры данных.
E-mail: kamali.gupta@chitkara.edu.in