Smart City Insights

www.sci.reapress.com

Smart City Ins. Vol. 1, No. 1 (2024) 63–72.

Paper Type: Original Article

Economic Impacts of Load Balancing in Cloud Computing

Arka Ghosh[1,*], Aryan Raj[1], Ananya Rout[1], Aparajita Dash[1], Swosti Priya Jena[1]

[1] KIIT Deemed to be University, Bhubaneswar, Odisha, India. 21052484@kiit.ac.in, 21052487@kiit.ac.in, 22057009@kiit.ac.in, 22057016@kiit.ac.in, 22057076@kiit.ac.in

Citation:

Received: 20 April 2024 *Revised: 13 June 2024* *Accepted: 11 August 2024*	Ghosh, A., Raj, A., Rout, A., Dash, A., and Jena, S. (2024). Economic Impacts of Load Balancing in Cloud Computing. *Smart City Insights, 1* (1), 63-72.

Abstract

This paper examines the economic implications of load balancing within various computing environments. Load balancing, a fundamental strategy in optimizing resource utilization, contributes significantly to cost savings, operational efficiency, and overall economic performance. By evenly distributing workloads across the various computing resources, load balancing minimizes idle time, maximizes resource utilization rates, and reduces the need for additional hardware investments. It leads to straightforward cost savings for businesses by optimizing existing infrastructure and minimizing operational expenses. Moreover, load balancing enhances scalability and flexibility, enabling most businesses to adapt to fluctuating demand patterns without significant infrastructure costs. Additionally, load balancing improves the reliability and resilience of IT systems, mitigating risks associated with system failures or future cyber-attacks. Overall, understanding the economic impacts of load balancing is crucial for businesses seeking to optimize their IT investments, improve operational efficiency, decrease system failures and gain a competitive edge in today's digital economy.

Keywords: Load balancing, Resource optimization, Cost savings, Economic impact, Cloud computing, Distributed systems, Workload distribution, Performance optimization.

1 | Introduction

Load balancing is the technique that distributes network traffic equally within multiple servers supporting an application. Economic factors can affect the load-balancing process in many ways, such as efficiency duration, overall system performance, and availability of resources. Let's assume a case where a high demand and load balancing algorithm may prioritize resource allocation based on economic consideration to optimize cost-effectiveness [1]. Load balancing is the process that gives directions to application servers and their clients.

Corresponding Author: 21052484@kiit.ac.in

Load balancing defines the proper distribution of the network across a group of servers, also known as a server farm or server pool. In cloud computing, load balancing can be implemented on various levels, such as the network, application, and database layers. Load balancing distributes the load among several nodes to improve the cloud service user experience without interfering with basic infrastructure [2]. Concerning load balancing techniques currently employed in cloud computing, the following metrics are mentioned:

Scalability

An algorithm can perform load balancing for a system with any finite number of nodes [3].

Performance assessment

Gauges the system's efficiency, aiming to streamline tasks within a reasonable budget, reducing task response times while maintaining acceptable delays [4].

Response time

Measures the efficiency of a load balancing algorithm within a distributed system, aiming to minimize delays [2]. Overhead assessment evaluates the associated overhead to ensure efficient load-balancing operations. Inefficient databases have a negative impact on system response times, performance, and, most importantly, outages [3]. The variety of cloud components and user demands causes some nodes to become overwhelmed when scheduling cloud resources, which impacts the utilization of the entire system and its resources [4]. Cloud computing enables resource sharing, handles enormous volumes of data processing, and offers limitless storage capacity. Currently, internet firms make big financial rewards using large amounts of user data to forecast future user behaviour and extract useful insights. Storage techniques are anticipated to evolve, moving away from the existing practice of keeping data on servers and personal PCs and towards the cloud.

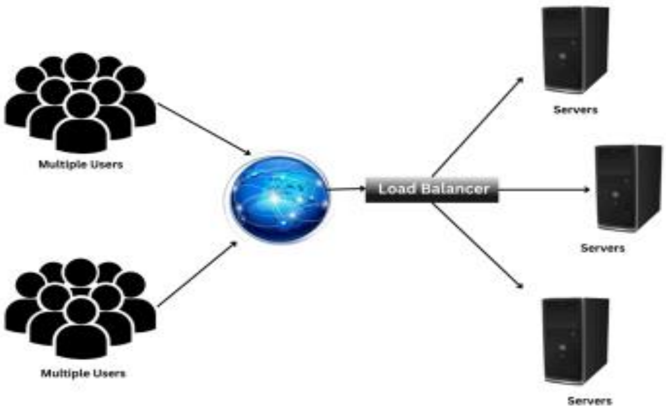

Fig. 1. Figure to illustrate Load balancing with multiple servers [5].

This change represents a paradigm shift in data storage methods, giving organizations operating in the digital sphere better access, scalability, and collaborative processing capabilities. Effective resource management reduces the risk of underutilized resources and helps organizations avoid needless costs associated with overprovisioning. So, load balancing helps distribute the workload evenly among all the servers [6]. This ensures that the server or a group of servers does not have a huge number of disproportionate requests, which helps the server reduce or crash down at any time.

Economically, it is beneficial as it helps the server deliver the content faster, be flexible while still being available at all times, and be cost-efficient. The traditional distributional network has undergone several changes in a few years. This improvement is environmentally friendly, reliable and controllable [7]. This paper tells us about the load balancing techniques [6]. The advantages and limitations of existing methods are highlighted with major challenges and develop efficient load-balancing algorithms in future. An edge computing model is the most efficient if its resources are used efficiently. For high efficiency, we need proper edge management resources [8]. Strong resource scheduling, efficient allocations, and powerful scalability techniques are required for resource management. Virtualization of software and hardware components is done in Virtual Machines (VMs), which are required to provide customer resources.

While submitting user requests, some VMs experience heavy traffic on user tasks, and some VMs experience low traffic. Therefore, edge service providers must deal with unbalanced machines with large user tasks and resource usage differences. Load balancing is used to redistribute the workload in a distributed system such as edge computing [9]. Load balancing accelerates response time, runtime, and system stability to improve performance [7]. A greater number of risk variables, such as the imbalance of national policies and economic structure, endanger economic security in the current period of economic globalization and liberalization. Due to economic security problems, monitoring is typically chosen alternatively based on certain criteria for monitoring economic crises. Network cloud is used so that cloud computing breaks down large data processing programs into several smaller ones [10]. These smaller programmers are then processed and analyzed by a system comprising numerous servers, which provides the user with the findings.

Additionally, transforming services offers fresh concepts for how information systems should function. Let's picture the frantic, ever-changing world of technology. Think of the need to balance loads as there is a huge load on the servers. As companies depend more on tech, managing resources well and keeping things running smoothly are key. All companies, mechanical, electrical, etc., are directly or indirectly based on tech. Therefore, managing servers and resources and keeping everything working smoothly is a key to becoming a good company. Load balancing is like a smart distributor. It spreads out work tasks over lots of computer resources [11].

This way, no single resource gets too overloaded. So, load balancing is a techy concept and helps save money and make the most of resources [1]. Efficient load balancing can enable greater flexibility and scalability, enabling organizations to achieve greater heights and save costs. Organizations can better respond to changing business dynamics and adapt to evolving business needs by allocating resources based on changing demand patterns. This flexibility is especially valuable in industries with time-sensitive or unpredictable requirements, where the ability to scale up or down quickly can mean the difference between success and failure. Thus, proper load balancing for adopting new tech, such as cloud edge computing, develops. These new techs offer more opportunities for better performance and lower costs. These new technologies enable organizations to achieve hybrid performance and cut down on costs. Thus, large computing resources can be accessed quickly without additional work. The bottom line is that proper load balancing is key to optimizing resource utilization, increasing productivity and saving money in many projects. Forward-thinking weight balancing makes companies more efficient, reliable, and productive. It gives them an edge in a fast-paced business world. A suitable study technique was used to investigate the load-balancing process and determine what causes load-unbalancing problems [8].

The literature review complied with the general research strategy, which describes the technique adopted to solve the load-unbalancing problem and defines its theories, methods, algorithms, approaches, and paradigms. The Constructive Generic Framework (CGF) methodology, which divides the load unbalancing problem into smaller processes, or the variables, factors, and parameters related to load balancing, was followed in the study of this subject. The literature review was further improved by adhering to Kitchenham's research guidelines for Systematic Literature Reviews (SLR), emphasizing cloud load balancing mechanisms studies. An SLR is an iterative research technique that other researchers might use to investigate further topics. Load balancing is like the work done by a manager in a restaurant. And let's assume a restaurant with six waitpersons. If customers were allowed to choose their waitpeople, 1 or 2 waiters could be overloaded with work while the others would be idle. To avoid this problem, the restaurant manager will assign the customers to specific waitpeople who are best suited to serve them according to their needs. So, companies follow the same process, with their applications running on multiple servers. When the user requests the application, the company first goes to the load balancer. Then, the load balancer will request each route of the single server in the server farm that is best suited to handle the request and fulfil the user's needs, just like the restaurant managers do. Load-balancing algorithms have two main categories.

Static load balancing

Static load balancing algorithms use the method in which the fixed rules are there, and the current server state has independence. The following are the examples of static load balancing method:

I. Round-robin method

II. Weighted round-robin method

III. IP hash method

Dynamic load balancing

Dynamic load balancing algorithms in which the examination of the current state of the servers is done before distributing traffic. The following are examples of dynamic load-balancing algorithms:

I. Least connection method

II. Weighted least connection method

III. Least response time method

IV. Resource-based method

Additionally, the research offers a fresh viewpoint on cloud computing load balancing. Researchers concentrating on job scheduling, resource scheduling, resource allocation, and resource management employ various load-balancing techniques [9].

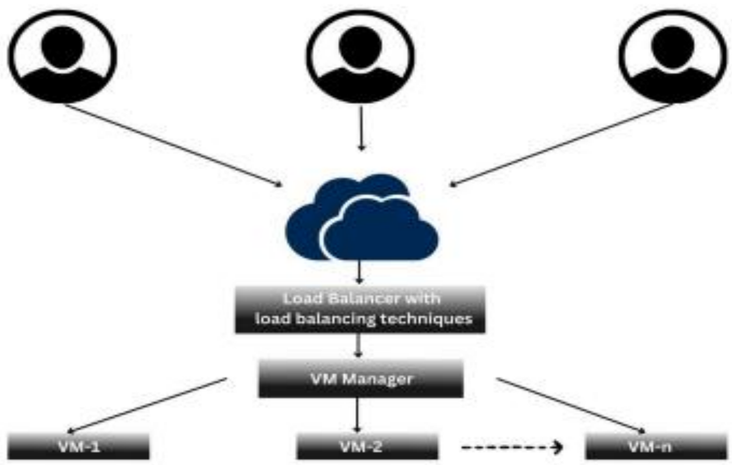

Fig. 2. Figure to illustrate Load Balancer with load balancing techniques.

2 | Literature Review

This paper presents a systematic and brief literature review of the existing load balancing techniques that are used or exist. This paper also reviews the advantages and disadvantages of several load balancing algorithms that are present and the important challenges of the algorithms which will be helpful for more efficient load balancing. The techniques can be developed for load balancing, and the main key points where future research can improve the load balancing process can be outlined. Many types of research have been done on cloud computing and its general challenges, including scheduling, resource provisioning, load balancing, etc. However, a little comprehensive part of the research is done on cloud load balancing. Load balancing algorithms are based on the system topology and availability of information about resources that are further categorized into two parts, i.e. static and dynamic. Static algorithms are the methods that don't depend on the system's current state; prior knowledge about the system is required, such as communication time, the processing power of system nodes, memory and storage devices capacity and many more. Static load balancing algorithms that use the round-robin approach. The main drawback of static load balancing algorithms is that the system's current state isn't considered while making the decisions, so it isn't a suitable approach. Dynamic algorithms are based on the current database of the system's state and are used to predict the unpredictable processing loads on a network storage virtualization basis [6].

L.D and Venkata Krishna [12] proposed a solution for load balancing, In which they stated that an algorithm named honey bee behaviour is solely inspired by load balancing to achieve a well-balanced load across VMs. The tasks that must be balanced are proposed as the BEES, and the VMs, the food sources with low-load machines, are represented as bees' destinations. A task loading to a VM is represented by a bee foraging a food source. According to the load and availability of the VMs, this update gives the user a clear idea of which task should be assigned to which VM. This method improves the overall throughput and reduces the time a task has to wait in a queue for the VM [4]. Thus, it reduced the response time and makespan. In cloud computing, we manage tons of data constantly moving around, and servers handle countless tasks to ensure they run smoothly.

A set of rules and guidelines are followed across servers to balance the task. However, Machine Learning (ML) will come in if we can teach the system to learn from past experiences and make smarter decisions in real-time. We want to teach the system to recognize the patterns in the data it sees. Some servers get busier, or a particular task requires more resources. Also, the system needs to be flexible enough to adapt to changes. Cloud is always an evolving sector with new tasks and servers going offline. So, our ML model has to be able to keep up and adjust its strategies on the fly. Finally, we need to be able to test and compare our ML model to check their actual work. We must develop fairways to measure their performance and check other old models. Efficient load balancing carries significant economic implications by enhancing resource utilization and reducing operational costs. When we evenly distribute workloads across servers or resources, various organizations try to optimize their infrastructure, which leads to cutting costs and capacity or some additional investments in hardware. It saves costs through lower energy consumption, minimized maintenance expenses, and extended hardware lifecycles. Moreover, streamlined operations foster higher productivity levels, enabling businesses to allocate resources more effectively towards innovation and growth initiatives.

Focusing on the literature review, the previous studies have explored all the aspects related to Technical problems of load balancing, focusing on algorithms, performance metrics and system-related architectures. Future research could explore the long-term economic benefits, incorporating factors such as scalability, adaptability to dynamic workloads, and the influence of emerging technologies like cloud computing and edge computing on cost optimization and resource utilization. While existing literature extensively covers technical aspects such as algorithm design and system performance, a scarcity of research delves into these strategies' economic ramifications and cost-effectiveness [3]. This paper presents a review that covers some significant concerns that the cloud load balancing needs rigorous attention but has not received enough attention in the technical or survey literature. As a result, we talk about some open research in this section. An algorithm's complexity is crucial in deciding how well any load-balancing technique works. Of the 35 possible technical papers examined for this study, only 7 define the associated algorithmic difficulty, which is 20%, and 28 do not define the algorithmic complexity, which is 80%. Due to the observation that most publications do not address algorithmic complexity, it is recommended that future researchers consider. Migration is always involved in a reactive load-balancing strategy, specifically task migration. Task migration always comes with a price tag known as the migration cost. The analysis shows a shortage of literature on migration costs in cloud load balancing besides Service level violations, task rejection ratio, and power consumption. It can be seen as a crucial avenue for academics working on reactive techniques with low migration costs in the future [10].

An analysis of load-balancing strategies in peer-reviewed publications is presented in this paper. The issue of load unbalancing in cloud computing and the contributing reasons to this problem were examined. We quickly reviewed the steps required in the load-balancing process and an abstracted load-balancing model. Following the rules of the CGF and the SLR methodology, an appropriate research approach was used to study the topic. At work, we posed and addressed a number of problem-related concerns. Several resources helped with the data search procedure. After a thorough review, it was determined that many load-balancing process issues remain unresolved. These issues can be resolved by implementing an effective and sophisticated load-balancing algorithm, particularly when evaluating algorithm complexity and additional QoS metrics. Additionally, the survey provides a taxonomy of certain algorithms that can help future researchers efficiently address load unbalancing problems. These include algorithms drawn from mathematics, ML, and natural phenomena.

3 | Proposed Model

There is a research gap in understanding the economic impact of load balancing. While it is widely recognized that load balancing can lead to cost savings and improved system performance, comprehensive studies do not quantify these benefits. Researchers could explore this area by analyzing the impact of load balancing on resource utilization, system uptime, and the need for additional hardware in different industries and use cases [7]. The biggest challenge is to gather the right kind of data, which is important for making decisions. We must ensure the system can handle big data and make decisions quickly without slowing everything down. Here, we need to take care of the cost involved. We need to find a balance between using resources cost-efficiently and load balancing. A notable research gap lies in the lack of comprehensive studies addressing the nuanced economic implications of load-balancing strategies. There's a significant hole in load-balancing research [5]. We're missing in-depth studies on how different load-balancing methods affect the economy. While existing literature extensively covers technical aspects such as algorithm design and system performance, there's a scarcity of research that delves into these strategies' economic ramifications and cost-effectiveness.

Moreover, research should investigate the scalability and adaptability of load-balancing solutions to dynamic workloads and changing business environments. It's about how load-balancing plans change with time and every day. It deals with shifting needs, uses resources to the max, and lowers operating costs [6]. Overall, bridging the gap between technical innovation and economic analysis is essential for informing strategic decision-making and maximizing the value of load-balancing initiatives in various organizational contexts. However, gaps exist concerning the economic impacts and financial considerations associated with load-balancing implementation. Many studies dive deeper into the cost-benefit analysis or provide various frameworks for assessing the economic capability of load-balancing strategies.

3.1 | Proposed Solution

The proposed solution is to achieve efficient load balancing and realize its economic benefits; organizations can implement a combination of load-balancing algorithms, automation tools, and monitoring solutions. Load balancing is generally balanced with many algorithms such as round-robin, least connections, or weighted least connections that can be used to distribute traffic or loads and tasks across multiple servers based on predefined criteria set by the organization [5]. Automation tools enable organizations to automate the work and scale resources based on workload demand, reducing manual work handling and improving operational efficiency. Monitoring solutions provide real-time insights into system performance, allowing organizations to identify bottlenecks and optimize resource allocation for maximum efficiency. When a user makes a request, the load balancer then requests a given server, and this method is repeated for each request. Load balancers' work defines which server should handle each request based on several algorithms. These algorithms have been divided into two main categories, i.e. static and dynamic.

Static load balancing algorithms

In static load balancing algorithms, the workload is distributed without considering the system's current situation. A static load balancer is unaware of which servers perform slowly and which are not used enough for proper response. It is usually assigned to the workload and the predetermined plan. The static load balancing method has a quick setup, but the results are inefficient. Round Robin DNS and client-side random load balancing are two common examples of static load balancing.

Dynamic load balancing algorithms

The Dynamic load balancing algorithms consider each server's current availability, workload, and health. Unlike the static load balancing algorithms, this algorithm can shift traffic from overburdened or poorly performing servers to underutilized servers, which helps keep the distribution even and efficient. This method uses several factors, such as server availability, the health and overall capacity of each server, the size of the tasks being distributed, and so on.

Round robin load balancing

It diverts incoming requests to a group of servers in a circular sequence. It is simple and easy to implement and may not consider server load or performance.

Weighted round Robin

It assigns different weights to servers based on their capacity or performance. It also distributes the load proportionally to the assigned weights and is useful when servers have different capacities.

Least connections load balancing

It helps direct traffic to the server with the fewest active connections. It helps to distribute load based on current server loads and may require continuous monitoring of server connections.

Weighted least connections

It is similar to a weighted round-robin but based on the number of active connections. Servers with higher weights get more connections, helping to balance the load based on capacity.

Least response time load balancing

It helps route traffic to servers with the lowest response times. It requires continuous server performance monitoring and ensures that requests are directed to the fastest and most responsive servers.

Adaptive load balancing

This load balancing can dynamically adjust to changing server loads and conditions. It uses real-time monitoring and feedback to make load-balancing decisions and provides flexibility in handling varying workloads.

Content-based load balancing

It considers the type or content of requests when distributing load and is useful when different servers have different capabilities for handling specific types of content.

Global Server Load Balancing (GSLB)

Balances loads across multiple data centres or geographically distributed servers and considers server health, proximity, and site traffic factors.

Queue-based load balancing

It distributes requests based on the server's queue length and helps prevent server overload by considering the server's ability to handle incoming requests [11].

Balancing simultaneous requests is highly complicated, non-trivial, and critical at times, which forces the addition of a greater number of systems or an external service to handle requests and balance resource utilization. Load balancing helps improve the system's performance by optimizing the use of resources, maximizing the throughput, and reducing latency. ML techniques have been increasingly used in load balancing to get optimal resource allocation and to improve the system performance in multiple domains, especially in distributed computing environments such as cloud computing and edge computing, as well as data centres.

Predictive load balancing: ML models are trained to predict future workload patterns based on the past data provided. These predictions can help preemptively reallocate resources to prevent overloading or underutilization of servers.

Dynamic resource allocation: ML algorithms can dynamically allocate resources based on real-time workload demands. For instance, reinforcement learning algorithms can learn optimal resource allocation policies by continuously observing system states and adjusting resource allocations to optimize performance metrics such as response time, throughput, or energy consumption.

Anomaly detection and fault tolerance: ML techniques can detect anomalies in system behaviour that may indicate impending failures or performance degradation. By identifying these anomalies early, load balancers can take proactive measures such as redistributing workloads or migrating services to healthier nodes to prevent system failures or service disruptions.

Quality of Service (QoS) optimization: ML algorithms can optimize load balancing strategies to meet specific QoS requirements such as response time, throughput, and availability. ML-based load balancers can dynamically adjust their strategy to prioritize critical workloads and ensure Service Level Agreement (SLA) compliance by continuously monitoring system performance metrics and user demands.

Traffic prediction and routing: ML models can analyze network traffic patterns to predict future traffic flows and optimize routing decisions. By considering network latency, bandwidth availability, and server capacities, ML-based load balancers can route incoming requests to the most suitable servers to minimize response times and maximize throughput.

Auto-scaling: ML-driven mechanisms can automatically provision or de-provision resources based on workload fluctuations. By analyzing historical workload data and predicting future demand, auto-scaling systems can scale infrastructure resources up or down to maintain optimal performance and resource utilization levels while minimizing costs. ML algorithms continuously learn from past experiences to improve load-balancing decisions. As we refine our skills, these systems adapt and evolve, becoming more adept at managing workloads and delivering superior performance.

3.2 | Analysis and Discussion

The economic impacts of load balancing can be analyzed from multiple perspectives, including cost savings, revenue generation, resource utilization efficiency, and customer satisfaction.

Cost savings: load balancing is crucial in reducing business operational costs. By distributing workloads evenly across servers or computing resources, organizations can prevent the overloading of specific resources, thereby avoiding costly downtimes and performance degradation. Efficient load balancing mechanisms help optimize resource utilization, reducing the need for additional hardware investments. It translates to direct cost savings for businesses, as they can achieve better performance with existing infrastructure.

Resource utilization efficiency: load balancing ensures that computing resources are utilized efficiently, which leads to higher productivity and cost-effectiveness. By dynamically allocating tasks to underutilized resources, load balancing minimizes idle time and maximizes resource utilization rates. This efficiency gains significance in cloud computing environments, where resources are provisioned and billed based on usage. Effective load-balancing strategies enable organizations to maximize their cloud investments by scaling resources according to demand and optimizing costs [11].

Scalability and flexibility: load balancing facilitates scalability by enabling seamless expansion or contraction of computing resources based on workload demands. This scalability ensures businesses can adapt to fluctuating demand patterns without incurring significant infrastructure costs. Moreover, load balancing enhances the flexibility of IT infrastructure, allowing organizations to deploy new services or applications more rapidly. This agility is crucial for businesses to stay competitive in dynamic markets and capitalize on emerging opportunities.

Revenue generation: besides cost savings, load balancing can directly contribute to business revenue generation. Load balancing enhances customer satisfaction and loyalty by ensuring high availability and responsiveness of services. Improved user experience leads to higher customer retention rates and increased revenue streams. Furthermore, by optimizing the performance of e-commerce platforms or online services, load balancing can boost sales conversions and transaction volumes, directly impacting the bottom line.

Risk mitigation: load balancing is vital in mitigating risks associated with system failures, cyber-attacks, or unexpected traffic spikes. By distributing workloads across multiple servers or data centres, load balancing minimizes the impact of individual failures or security breaches. This resilience enhances business continuity and reduces the likelihood of revenue loss due to downtime or service disruptions. Moreover, load balancing enables proactive capacity planning and traffic management, helping organizations anticipate and address potential risks before they escalate [9].

4 | Conclusion

This paper presents a comparative and proper study of load balancing and the approaches from the reviewed articles. The problem faced in the load unbalancing process in cloud computing and the factors that lead to the problem are discussed. A proper research methodology was followed to solve the problem with the study guidelines. By distributing the network traffic uniformly, the load balancing enables the authorities to avoid resource failure using resource overload. Analyzing the economic impacts of load balancing involves many factors, such as operational efficiency, infrastructure costs, and revenue generation.

The economic impact of load balancing is usually positive, as it saves the cost, improves performance, and generates revenue. However, the specific implications will differ accordingly by factors such as the scale of operations, industry, and the efficiency of the load-balancing implementation. Businesses must assess their requirements carefully and evaluate the potential return on the investment before implementing load-balancing solutions. As we know, the web server of a system uses several load-balancing methods to distribute its load among available web resources. The resources used are getting expensive day by day. Therefore, efficient cost optimization mechanisms support small organizations and the industry. A detailed description of the existing approaches, strengths, limitations, and future scope has been analyzed, and an adequate radiance has been thrown to these techniques. This paper investigates the problems of load-balancing algorithms to propose future load-balancing methods that will be even more effective. Most assessed articles did not consider significant and essential QoS measures for analysis. In-depth discussions of a few crucial QoS indicators, such as migration time, cost, power consumption, service level violation, task rejection ratio, and degree of balance, are absent from the examined literature.

Additionally, our research showed that algorithm complexity is not considered when assessing the load-balancing method's performance. As a result, 80% of the workers do not believe it. Additionally, 94.44% of all load-balancing techniques have been applied on simulator platforms. After a thorough review, it was determined that many load-balancing process issues remain unresolved. These issues can be resolved by implementing an effective and sophisticated load-balancing algorithm, particularly when evaluating algorithm complexity and additional QoS metrics. The conclusion and results of load balancing typically depend on the specific context and goals of the implementation. However, effective load-balancing strategies generally aim to optimize resource utilization, improve system performance, enhance reliability, and ensure equitable distribution of workloads across servers or resources.

Author Contributions

conceptualization, S.P. and A.A.; methodology, A.A.; software, S.M.and A. S; validation, M.M., A.P. and M.M.; formal analysis, A.A.;writing-original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability

All the data are available in this paper.

Conflicts of Interest

The authordeclare no conflict of interest.

References

[1] Ke, W., Wang, Y., & Ye, M. (2020). GRSA: service-aware flow scheduling for cloud storage datacenter networks. *China communications*, *17*(6), 164–179. DOI:10.23919/JCC.2020.06.014

[2] Mohapatra, H., & Rath, A. K. (2020). Fault-tolerant mechanism for wireless sensor network. *IET wireless sensor systems*, *10*(1), 23–30. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-wss.2019.0106

[3] Pérez, F. M., Fonseca, I. L., Martínez, J. V. B., & Maciá-Fiteni, A. (2023). Distributed architecture for an elderly accompaniment service based on IoT devices, AI, and cloud services. *IEEE multimedia*, *30*(1), 17–27. DOI:10.1109/MMUL.2022.3206398

[4] Mohapatra, H., & Rath, A. K. (2019). Fault tolerance in WSN through Pe-Leach protocol. *IET wireless sensor systems*, *9*(6), 358–365. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-wss.2018.5229

[5] Luo, S., Xing, H., & Fan, P. (2021). Softwarized IP multicast in the cloud. *IEEE network*, *35*(6), 233–239. DOI:10.1109/MNET.100.2100045

[6] Mohapatra, H., & Rath, A. K. (2019). Detection and avoidance of water loss through municipality taps in India by using smart taps and ICT. *IET wireless sensor systems*, *9*(6), 447–457. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-wss.2019.0081

[7] Zhou, G., Tian, W., & Buyya, R. (2023). LPT-one and bfd-one search algorithms for load balance and bin-packing of cloud computing. *2023 IEEE 9th international conference on cloud computing and intelligent systems (CCIS)* (pp. 521–525). IEEE. DOI: 10.1109/CCIS59572.2023.10263040

[8] Mohapatra, H., & Rath, A. K. (2020). Survey on fault tolerance-based clustering evolution in WSN. *IET networks*, *9*(4), 145–155. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-net.2019.0155

[9] Fan, W., He, J., Han, Z., Li, P., & Wang, R. (2020). Intelligent resource scheduling based on locality principle in data center networks. *IEEE communications magazine*, *58*(10), 94–100. DOI:10.1109/MCOM.001.1900324

[10] Mohapatra, H., & Rath, A. K. (2021). Fault tolerance in WSN through uniform load distribution function. *International journal of sensors wireless communications and control*, *11*(4), 385–394. https://scholar.google.com/citations?user=hZquqF8AAAAJ&hl=en&oi=sra

[11] Deng, Y., Chen, X., Zhu, G., Fang, Y., Chen, Z., & Deng, X. (2022). Actions at the edge: jointly optimizing the resources in multi-access edge computing. *IEEE wireless communications*, *29*(2), 192–198. DOI:10.1109/MWC.006.2100699

[12] L.D., D. B., & Venkata Krishna, P. (2013). Honey bee behavior inspired load balancing of tasks in cloud computing environments. *Applied soft computing*, *13*(5), 2292–2303. https://www.sciencedirect.com/science/article/pii/S1568494613000446