IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Load Balancing using Weight Based scheme in AWS

Bhavana MK, Meghana HM, Poornima RM
Department of Information Science and Engineering
Global Academy of Technology Bangalore, India
bhavana1ga20is025@gmail.com, meghana1ga20is168@gmail.com, poornimarm.gat@gmail.com

Abstract: *This paper presents research on load balancing methods for cloud computing platforms. Load balancing is an important element of cloud systems since it allows for optimal resource utilization and consistent performance levels. The study investigates several load balancing technologies and evaluates their effectiveness in coping with the dynamic and heterogeneous nature of cloud workloads. The study conducts experiments and analyses to discover significant parameters influencing load balancing performance, such as workload allocation, server capacity, and network latency. With the weight-based technique, different servers are dynamically given requests based on their weights, which are determined by the processing capacities of the servers. By distributing the work proportionately, this approach aims to decrease user request response times and avoid server overload. This paper presents an optimized load distribution approach that raises the efficiency and stability of cloud computing systems. The findings enable to improve load balancing methods matched to the particular requirements and issues of cloud computing, eventually enhancing system scalability and reliability*

Keywords: AWS, Load balancing, Weight schemes

I. INTRODUCTION

Load balancing is a crucial cloud computing technique that ensures efficient distribution of workloads across multiple resources. Amazon Web Services (AWS) offers several load balancing schemes, including weight-based schemes. These models use weights to distribute traffic evenly among available load balancer instances. Load balancing is the technique of distributing network load among system resources. This leads to faster job response times and more efficient resource utilization. This method prevents overloading and under loading of a small number of hubs. Redundancy allows for improved dependability and data availability by distributing load among multiple components.
The system calculates load based on memory, CPU, and network loads.
Load balancing seeks to meet two primary objectives:

1. High availability: Load balancing distributes traffic over numerous servers, ensuring that the failure of one server does not affect the application's overall availability. If one server fails, the load balancer routes traffic to other available servers, reducing downtime and enhancing overall performance.
2. Resource Optimization: Load balancing enables managers to balance network traffic between servers, ensuring that each server receives an equal proportion of requests and burden. Load balancing helps to minimize congestion and maximize resource utilization, resulting in more efficient server usage and improved application performance.

Weight-based load balancing guarantees that traffic is evenly distributed among available instances. This prevents overloading a single instance and maximizes resource utilization. By giving weights to distinct instances, weight-based load balancing allows for traffic control. For example, if an instance has performance concerns or is not operational, the weight allocated to it might be lowered, reducing traffic to that instance. AWS gives weights to instances depending on a variety of factors, including instance type and availability zone. This eliminates the need to manually adjust weights, minimizing complexity and mistake. Weight-based load balancing offers information on the performance of individual instances. Administrators can increase performance by monitoring instance weights.
The rest of the section is structured as follows:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

In section 2, it consists of the brief literature review. In section 3, it consists of table of comparison results. section 4 provides conclusion and references

II. LITERATURE REVIEW

The aim of this literature review is to examine the numerous aspects and perspectives associated with load balancing in AWS.

The study [1] specializes in resource scheduling and load balancing methods for cloud computing. It examines and evaluates various algorithms developed for optimal resource allocation and task distribution in cloud systems. The authors explore the characteristics, advantages, and limitations of different scheduling and load balancing strategies, with the goal of providing insights into their application and performance in various cloud computing scenarios. Overall, the research adds to our understanding of successful resource management tactics in cloud environments, which may aid in the design and implementation of optimal cloud systems.

This research [2] explores a novel way to load balancing in cloud computing using the Inspired Lion Optimization Algorithm. It will most likely explore the Lion Optimization Algorithm's core ideas and how they are applied to the issues of load balancing in cloud systems. The recommended method is likely to focus on optimizing resource utilization, reducing reaction times, and increasing overall system efficiency. The authors are likely to share insights into the design and implementation specifics of their methodology, demonstrating its potential advantages over traditional load balancing methods. The paper's empirical evaluations or simulations are anticipated to demonstrate the Inspired Lion Optimization Algorithm's performance and effectiveness in improving load balancing capabilities within cloud computing infrastructures. Overall, the study helps to further load balancing approaches in cloud computing leading to more resilient and scalable cloud-based systems

This paper [3] A weight-based technique is developed to address load balancing issues in the cloud computing environment. It is likely to implement a system for assigning weights to resources based on factors such as processing power and current workload. The weights are then applied to a load balancing algorithm, which optimises the distribution of inbound requests or tasks among available resources. The article is likely to include empirical assessments or simulations to demonstrate the usefulness of a weight-based system in increasing resource utilisation and improving system performance when compared to standard load balancing methods. As a whole, the goal of this work is to give a viable approach for tackling load balancing difficulties in cloud environments, which can increase the scalability and performance of cloud-based

The study [4] paper seeks to improve disc load distribution, optimize resource utilization, and reduce response times in cloud-based storage infrastructures. The study most likely provides empirical evaluations or simulations to demonstrate the effectiveness of the proposed strategy in improving disc load balancing compared to existing ways, contributing to developments in cloud computing resource management. They introduce an updated Lion optimization technique in conjunction with the min-max strategy to achieve more effective resource allocation and workload distribution across disc storage systems.

This paper [5] provides a systematic framework for classifying load balancing options for cloud services. These technologies will most likely be classified in a hierarchical order depending on features like functionality, architecture, and implementation. The article will most likely investigate the taxonomy's various divisions and subcategories, as well as provide a comprehensive overview of the load balancing landscape in cloud systems. The authors anticipate that this classification will help researchers, practitioners, and stakeholders better understand the various load balancing solutions. A hierarchical taxonomy is intended to shed light on the similarities, differences, and relationships between various load balancing systems, enabling for the selection and implementation of appropriate solutions for certain cloud computing scenarios.

The study[6] studies how to optimise server placement and task allocation in edge computing networks to achieve effective load balancing. It will most likely investigate methods for evaluating the optimal location of servers near edge devices and users in order to reduce latency and improve response times. The article will most likely discuss approaches and methodologies for assigning workloads to edge servers based on workload characteristics and network constraints. The project will use empirical assessments or simulations to examine the performance of various server placement and work distribution strategies in edge computing environments. Overall, the study contributes to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

improving the efficiency and scalability of edge computing networks by addressing load balancing difficulties through optimal server location and task distribution strategies.

This paper [7] provides a load balancing technique for data centers that optimizes cloud computing applications. It is likely to propose a novel way for efficiently spreading workload across servers within data centers, with the goal of improving the performance and scalability of cloud-based applications. The method may take into account server capacity, current demand, and network conditions in order to dynamically assign resources and balance the load efficiently. The authors are likely to undertake empirical evaluations or simulations to evaluate the suggested algorithm's usefulness and efficiency when compared to existing methods. The paper advances load balancing strategies designed for cloud computing settings by providing practical solutions to improve the overall performance and stability of cloud-based applications.

The primary objective of the paper [8] is to show that it provides a thorough investigation and comparison of several types of load balancers. It investigates various load balancing approaches, including round-robin, least connections, IP hash, and weighted round-robin, among others. The authors analyze various approaches using factors such as performance metrics, scalability, fault tolerance, and ease of deployment. The study also offers empirical evaluations or simulations to determine the efficacy and efficiency of each load balancer type in various cloud computing environments. Overall, the study intends to provide insights into the merits and disadvantages of various load balancing systems, assisting cloud practitioners in determining the most effective strategy for their individual deployment goals.

The research [9] conducts a meta-analysis of load balancing and server consolidation in cloud computing environments. It synthesizes insights from previous research projects by examining various load balancing and server consolidation approaches used in cloud systems. The authors explore the effectiveness, efficiency, and issues associated with various methodologies, with the goal of providing a thorough review of current load balancing and server consolidation practices. The meta-study is likely to provide insights into developing trends, identify gaps in present research, and make recommendations for future studies, all of which will assist in advance load balancing and server consolidation methods in cloud computing.

The context presented in this research [10] is a survey of cloud computing and load balancing that is believed to provide a comprehensive overview of existing literature, research, and practices in the field. It explores various aspects of cloud computing such as its architecture, service models and deployment strategies. In addition, the study will likely explore the challenges and opportunities associated with load balancing in cloud environments, including issues related to scaling, resource allocation, and performance optimization. The authors can discuss various load balancing techniques and algorithms used in cloud services, analyzing their effectiveness and suitability for different use cases. Additionally, the survey may highlight emerging trends, advances and future trends in cloud computing and load balancing research. Overall, the article aims to provide a comprehensive overview and analysis of the state of the art in cloud computing and load balancing, a valuable resource for researchers, practitioners and stakeholders in the field.

This paper [11] is to show that L3B is likely designed to efficiently distribute incoming requests or tasks among available resources in the cloud architecture at a low level, possibly closer to the network layer. The authors are likely to provide L3B's design and implementation details, emphasizing its usefulness in optimizing resource utilization and boosting overall system performance. L3B may use algorithms or heuristics designed specifically for low-level load balancing jobs, providing benefits in terms of scalability, fault tolerance, and responsiveness. The article will most likely include empirical assessments or case studies that demonstrate L3B's effectiveness and efficiency in real-world cloud computing scenarios. The study advances load balancing techniques by presenting a specialized solution designed for low-level activities in cloud systems.

This paper [12] focuses on highlights research into a load balancing solution designed exclusively for cloud computing environments. It analyses different load balancing approaches and algorithms in order to determine the most effective method for optimizing resource allocation and workload distribution in cloud systems. The authors explore the characteristics, benefits, and drawbacks of several load balancing strategies, evaluating their applicability for various cloud computing scenarios. The research will include empirical assessments or simulations to evaluate the performance of the proposed load balancing method to existing approaches. The study adds to the body of knowledge in cloud computing by providing insights into effective load balancing solutions, which may improve the efficiency and scalability of cloud-based services.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

This paper [13] Performs a comparison of static and dynamic load balancing approaches in cloud computing. Its purpose is to explore the characteristics, advantages, and disadvantages of both static and dynamic load balancing systems. The authors investigate how these algorithms perform in terms of resource utilisation, system responsiveness, and scalability across various workload scenarios. The paper compares and tests the performance of static and dynamic load balancing approaches in various cloud computing environments. Overall, the study aims to provide insights into the efficacy and usability of static and dynamic load balancing solutions, thereby supporting cloud practitioners in identifying the optimal option for their specific deployment requirements.

This paper [14] Examines the weighted round Robin (WRR) load balancing algorithm used in cloud-based web services. It will probably look at how WRR distributes incoming requests between servers according to defined weight values, with the aim of optimizing resource usage and improving system performance. Studies most likely use empirical estimates or simulations to determine the impact of WRR on characteristics such as response time, throughput, and server utilization in cloud environments. The purpose of the paper is likely to highlight the advantages and disadvantages of WRR in cloud-based web services, comparing it with other load balancing algorithms.The study improves specialized load balancing strategies for cloud-based applications and provides insights into optimization of web service delivery

III. TABLE OF COMPARISON RESULT

Table of Comparison Result

Sr No	Author /Year	Research work/Paper	Objective	Methodology	Technique	Achieved
1.	Vijay et al, 2023	Resource Scheduling and Load Balancing Algorithms in Cloud Computing	The goal of this study is to introduce a novel resource planning technique that combines Genetic technique (GA) and Particle Swarm Optimisation (PSO) in a cloud computing environment. The job's purpose is to optimise task assignment and scheduling in a computer system in order to maximise resource utilisation while minimising response time2.	The authors offer a new hybrid resource planning approach for cloud computing that optimises virtual machine (VM) allocation using a genetic algorithm (GA) and particle swarm optimisation (PSO) for cloud applications. The algorithm is divided into two phases: GA and PSO. Its goal is to minimise execution time, total cost, and duration while maximising resource utilisation and load balancing, outperforming existing algorithms.	Balancing and scheduling techniques in cloud computing, such as preemptive scheduling, SJF, FCFS, IP Hash, Least Connections, and Round Robin. It uses optimization techniques such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Floral Pollination Algorithm (FPA) to improve labor allocation and resource utilization.	The research describes a hybrid approach that combines genetic and particle swarm optimisation techniques for work sharing in cloud services. A CloudSim assessment compares its performance to existing approaches, concentrating on variables such as completion time, total cost, and duration to demonstrate its usefulness in increasing resource utilisation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

2.	Kaviarasan et al, 2023	Effective load balancing approach in cloud computing using Inspired Lion Optimization Algorithm	Current load balancing algorithms suffer from limitations in terms of time complexity, convergence, and security policy due to a lack of energy optimisation and dynamic adaptation. There is also a scarcity of bio-influenced metaheuristic algorithms that employ lion behaviour for effective cloud load balancing.	The authors create a bio-inspired algorithm based on lions' hunting and mating behaviour to handle load balancing in cloud environments. Their approach includes cloud node initialization, fitness evaluation, task identification, and handover between underutilised and overutilized nodes using the suggested Leon optimisation algorithm, which is explained using pseudocode and parameter explanations.	Bio-inspired metaheuristic algorithms, such as the Lion Optimisation Algorithm (LOA), use natural behaviour to solve complicated problems. The inspired Lion Optimisation Algorithm (ILOA), developed for cloud load balancing, examines node performance, separates nodes into groups, and employs transfer operators to maximise resource utilisation.	The research offers an algorithm, ILOA (Inspired Lion Optimisation Algorithm), that mimics lion behaviour in work assignment in clouds and outperforms existing metaheuristic algorithms. The review focuses on enhanced ILOA and detection, convergence, and task migration success, as well as improvements in cloud system performance, response time, migration, and failover.
3.	Uparosiya1 et al, 2023	Overcome load balancing problem by wight based schem eon cloud	The paper proposes a weight-based system to improve cloud service load balancing. The goal is to improve performance by optimizing task scheduling and reducing execution time, energy consumption, cost and transitions. It compares the proposed system with existing load balancing algorithms such as round robin, weighted round robin, centralized queue, min-min, bee foraging and genetic algorithm.	The article describes a weighted load balancing solution for cloud services that improves work scheduling, execution time, energy consumption, cost, and migration reduction. It compares this method to other load balancing algorithms such as round robin, weighted rework, centralised queue, min-min, bee foraging, and genetic algorithms algorithm.	The study describes a weight-based approach to cloud computing that improves task scheduling and load balancing while reducing execution time, energy consumption, cost, and migrations. It compares the proposed system against other load balancing algorithms, such as round robin, weighted round robin, centralised queue, min-min, bee foraging, and genetic algorithms.	The article proposes a weight-based method for increasing cloud service performance by optimising task scheduling and load balancing while lowering execution time, energy consumption, cost, and handovers. It compares this method to other load balancing algorithms, such as round robin, weighted round robin, centralised queue, min-min, bee foraging, and genetic algorithm.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

4.	Adaikalaraj et al,2023	To improve the performance on disk load balancing in a cloud environment using improved Lion optimization with min-max algorithm	The purpose of this paper is to offer an intelligent virtual machine scheduling model that uses machine learning techniques to improve disc load balancing performance in a cloud environment1. The study discusses two approaches: a min-max load balancing method and an advanced lion optimisation (ILO) strategy. According to the article, the proposed strategy outperforms current methods for data centre utilisation, virtual machine migration, energy consumption, and load balancing.	The virtual network, the methodology use the ILO algorithm in conjunction with a min-max load balancing approach. The study assesses data centre usage, virtual machine migration, energy consumption, and load balancing vs existing methods like random selection and machine learning.	The study uses Support Vector Regression (SVR) to model input-output interactions using a Gaussian kernel function to determine data point similarity. In addition, it presents the Min-Max optimisation approach and the ILO algorithm, which were inspired by lion hunting activity in space exploration.	The article describes an effective virtual machine allocation strategy with advanced disc load balancing in cloud environments that employs machine learning. It describes a sophisticated optimisation technique that employs a min-max approach to reduce network load, switching costs, latency, and power consumption.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

5.	Afzal et al,2021	Load balancing in cloud computing – A hierarchical taxonomical classification	The article presents a complete taxonomy of load balancing algorithms in cloud computing, outlining the elements that contribute to load imbalance as well as the obstacles of algorithm creation. It also proposes future research areas that will help to find acceptable solutions to the load imbalance problem in cloud computing.	The author follows Copilot's feedback guidelines, using markdown, code blocks, and LaTeX. They extract load balancing information from the Edge browser's online context, propose research questions, and provide a taxonomy-based classification of load balancing approaches, including advantages and disadvantages. There are challenges and lingering difficulties with cloud load balancing.	The authors use a systematic literature review (SLR) to collect and analyse data, followed by a constructive common framework (CGF) to address load imbalance iss ues. They present a taxonomy-based classification system for load balancing algorithms and conduct a comparative analysis of their advantages, disadvantages, challenges, and unanswered problems.	The goal is to investigate the factors that contribute to load imbalance in cloud computing, highlight the importance of load balancing solutions, systematically rank current approaches based on various criteria, and analyse the challenges associated with developing effective load balancing algorithms for cloud environments.
6.	HUANG et al, 2021	Server Placement and Task Allocation for Load Balancing in Edge-Computing Networks	The article proposes a novel technique to edge server placement and job allocation in edge computing networks, with the goal of reducing the most demanding workload among distance-weighted servers.	The authors approach server placement and job allocation as a mixed integer linear programming problem in order to reduce the maximum workload among Edge servers. They offer an integrated strategy based on the simulated annealing process (ESP) using Lagrangian duality theory and	The paper uses mixed integer linear programming (MILP) for server placement and work allocation, Lagrangian duality theory to simplify constrained optimisation, and the subgradient method for convex functions that are not differentiable. Simulated annealing is also used in stochastic optimisation to determine the optimal location of	This article provides a load balancing model that considers server load and transmission distance, as well as an integrated system that assigns and organises work on the computer using annealing simulation and Lagrangian duality theory. owner. Evaluation using K-medoids clustering and the greedy technique shows improved load

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

				the sub-hierarchical method, comparing it to K-medoids and greedy clustering on structures with changing network links, and investigating the impact of the distance.	edge servers.	balancing and decreased server effort across a range of topology sizes.
7.	SHAFIQ et al,2019	A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications	The study addresses difficulties such as incorrect task-resource mapping, unbalanced server workloads, and fluctuating task demands. It outlines a system for task scheduling and load balancing that prioritises quality of service. It aims to enhance the duration, execution time, cost of SLA violations, and resource utilisation in cloud systems.	The authors employ CloudSim to test their load balancing method in a cloud computing environment, and they provide a two-layer structure that prioritises job scheduling and resource allocation. QoS criteria drive task assignment to appropriate virtual machines, while migration strategies improve workload distribution. When compared to the dynamic LBA algorithm, performance is measured in terms of duration, execution time, and resource utilisation.	Virtualization optimises resource usage by operating several virtual computers on a single physical machine, whereas task scheduling allocates jobs to optimum performance. Load balancing distributes workload across virtual machines to improve system stability, using Dragonfly's optimisation method to establish ideal thresholds.	To balance VM loads, the technique considers critical QoS parameters such as task scheduling, resource deadlines and completion times, workload migration, and CPU reallocations. This improvement reduces latency and execution time while increasing resource utilisation, outperforming traditional load balancing methods in performance tests.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

8.	Simjanoska et al, 2019	Low Level Load Balancer in the Cloud	The study investigates cloud computing load balancing and server consolidation strategies, offers a taxonomy, evaluates existing solutions, and addresses issues before proposing future research areas.	This article compares five commercial load balancers (AVI Vantage, F5, Barracuda, A10, and Citrix BLX) based on their features, advantages, and disadvantages. The goal is to help IT organisations choose the most appropriate load balancer for their needs.	The document explains how to improve Citrix ADC BLX network performance via load balancing techniques including round robin and hashing, as well as the Data Layer Development Kit (DPDK). In addition, he is investigating Global Server Load Balancing (GSLB) to improve application availability and response time by sending consumers to the closest data centre based on their location.	The study evaluates five software load balancers (AVI Vantage, Barracuda, F5, A10, and Citrix BLX) based on their features, scalability, performance, compatibility, automation, and security. It assists IT organisations in selecting the finest load balancer for their organisations based on unique requirements and preferences.
9.	Ala'anzy et al, 2019	Load Balancing and Server Consolidation in Cloud Computing Environments: A Meta-Study	The goal of this study is to evaluate published load balancing techniques for server consolidation in cloud computing settings and give a taxonomy and meta-analysis of the most successful strategies. The report also discusses problems, measurements, and potential future research areas in the topic.	The methodology consists of a systematic literature review (SLR) that examines load balancing and server consolidation algorithms in cloud computing, using predetermined search criteria and quality assessment. The authors suggest a taxonomy based on common features, present descriptive statistics, and address problems and potential research areas.	The paper gives an overview of load balancing and server consolidation methods in cloud services, introduces a taxonomy, analyses current solutions, and reacts to difficulties and future research objectives.	The paper provides a comprehensive overview of cloud computing load balancing and server consolidation algorithms, including a proposed taxonomy, analysis of existing techniques and mapping of challenges and future research directions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

10.	Jyoti et al, 2019	Cloud computing and load balancing in cloud-computing survey	Propose a cloud-based platform for data science education that includes online learning and collaborative capabilities. Confirm its usefulness by doing a case study at an Indian institution and a data science course, comparing results and satisfaction to traditional classrooms.	collected data from multiple sites such as Twitter, GitHub, Stack Overflow, and Kaggle to analyse data scientists and engineers'behaviour patterns. Using Python, R, and Gephi, they conducted sentiment and network analyses to acquire insights and recommendations on industry difficulties and future directions.	Edge computing uses edge devices to process nearby data, reducing latency and cloud bandwidth, while fog computing extends this with intermediate cloud nodes to add additional capabilities. Multi-agent systems contain independent agents that cooperate to achieve goals, and reinforcement learning allows agents to learn by exploring and using optimal strategies without explicit instruction.	The Confluence Framework integrates edge and fog computing to achieve efficient computing, while an optimal algorithm enhances job allocation and scheduling to increase efficiency and scalability. Real-world evaluations have yielded remarkable results in fields as diverse as smart cities and healthcare.

IV. CONCLUSION

According to the literature survey the evaluation of Load Balancing for cloud systems concludes by showing how well it performs in terms of resource allocation optimization and service performance enhancement. The method's ability to account for factors such as service demand and server capacity enhances load distribution throughout the cloud infrastructure, hence enhancing overall system efficiency. The survey points out the need for greater research to handle dynamic workload fluctuations and scalability issues in cloud systems. When everything is taken into account, the findings demonstrate how important sophisticated load balancing algorithms are to maximizing resource utilization and ensuring excellent service delivery in cloud computing. The crucial function of load balancing strategies in cloud systems is to enhance system performance and maximizing resource use. A comprehensive overview of the topic is provided by the survey, which looks at several elements that impact load balancing, including resource availability and workload characteristics. Still, it is clear that there are research gaps concerning scalability and dynamic load balancing, which points to possible directions for further study. Further, the survey emphasizes how crucial it is to continue researching load balancing methods in order to keep up with the changing needs of cloud computing.

REFERENCES

[1]. Vijay R, Sree TR. Resource Scheduling and Load Balancing Algorithms in Cloud Computing. Procedia Computer Science. 2023 Jan 1;230:326-36.
[2]. Kaviarasan R, Balamurugan G, Kalaiyarasan R. Effective load balancing approach in cloud computing using Inspired Lion Optimization Algorithm. e-Prime-Advances in Electrical Engineering, Electronics and Energy. 2023 Dec 1;6:100326.
[3]. Mishra SK, Sahoo B, Parida PP. Load balancing in cloud computing: a big picture. Journal of King Saud University-Computer and Information Sciences. 2020 Feb 1;32(2):149-58.
[4]. Adaikalaraj JR, Chandrasekar C. To improve the performance on disk load balancing in a cloud environment using improved Lion optimization with min-max algorithm. Measurement: Sensors. 2023 Jun 14:100834.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

[5]. Afzal S, Kavitha G. Load balancing in cloud computing–A hierarchical taxonomical classification. Journal of Cloud Computing. 2019 Dec 23;8(1):22.

[6]. Huang PC, Chin TL, Chuang TY. Server placement and task allocation for load balancing in edge-computing networks. IEEE Access. 2021 Oct 4;9:138200-8.

[7]. Shafiq DA, Jhanjhi NZ, Abdullah A, Alzain MA. A load balancing algorithm for the data centres to optimize cloud computing applications. IEEE Access. 2021 Mar 10;9:41731-44.

[8]. Moharir M, Shobha G, Oppiliappan A, GVL RM, Pandit SN, Akash R, Saxena M. A study and comparison of various types of load balancers. In2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) 2020 Dec 1 (pp. 1-7). IEEE.

[9]. Ala'Anzy M, Othman M. Load balancing and server consolidation in cloud computing environments: a meta-study. IEEE Access. 2019 Sep 30;7:141868-87.

[10]. Jyoti A, Shrimali M, Mishra R. Cloud computing and load balancing in cloud computing-survey. In2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2019 Jan 10 (pp. 51-55). IEEE.

[11]. Simjanoska M, Ristov S, Velkoski G, Gusev M. L3B: Low level load balancer in the cloud. InEurocon 2013 2013 Jul 1 (pp. 250-257). IEEE.

[12]. Wang Q, Liu D. Research on load balancing method in cloud computing. In2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) 2018 Oct 12 (pp. 1489-1493). IEEE.

[13]. Deepa T, Cheelu D. A comparative study of static and dynamic load balancing algorithms in cloud computing. In2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) 2017 Aug 1 (pp. 3375-3378). IEEE.

[14]. Huang PC, Chin TL, Chuang TY. Server placement and task allocation for load balancing in edge-computing networks. IEEE Access. 2021 Oct 4;9:138200-8.