

Challenges of Load Balancing Techniques in Cloud Environment: A Review

Ola Hani Fathi Sultan

ola.20enp21@student.uomosul.edu.iq

Turkan Ahmed Khaleel

turkan@uomosul.edu.iq

Computer Engineering Department, College of Engineering, University of Mosul, Mosul, Iraq

Received: 26/5/2022

Accepted: 19/7/2022

ABSTRACT

To highlight issues and challenges of the Load Balancing Algorithm (LBA) in the context of cloud computing. The authors of this work discussed load balancing and the load balancing problems in cloud computing, how to solve them, the basic techniques used to maintain load over the cloud, and the benefits and drawbacks of these algorithms. Furthermore, the static and dynamic load balancing strategies are compared. Load balancing methods in the cloud for data centers must avoid VM overload/low load, which is a challenge in cloud computing. Load balancing mechanisms aid in the management of workloads. The primary goal of load balancing among machines is to spread workloads and computing resources across one or more servers while lowering energy consumption and maximizing resource usage. Cloud computing is a type of computing in which scalable resources are made available on-demand and are frequently virtual, as in an online service. For a variety of reasons, including cost saving, speed, efficiency, performance, and security, cloud computing has recently become a popular choice for individuals and enterprises. Cloud computing uses the Internet to deliver a variety of services to a company's computers and devices. These resources can include data storage, databases, servers, networks, and software, among other tools and applications. In this paper, developers and researchers must design and construct a load balancer for parallel and distributed cloud systems.

Keywords:

Cloud computing; Cloud data center; Load balancing; Load balancing algorithm; Load balancing challenges.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).
<https://rengj.mosuljournals.com>

1. INTRODUCTION

Cloud computing has arisen as a fascinating invention with the rapid advancement of computer technology allowing customers to use services at any time or on a leased basis[1]. Several businesses, often known as cloud providers (Google, Amazon, IBM, and others) help us provide cloud services to our clients. By appropriately distributing resources, the extreme center of cloud computing strives to achieve the greatest extreme system execution and efficiency[2]. In a typical cloud computing environment, there are two components: the front

side and the backside. The front side is on the user's side and is accessible via an internet connection, while the backend is concerned with the cloud service model. Cloud computing services can be accessed from a variety of different and widely distributed resources. On-demand services are provided by distributed computers[3]. Services include Software as a Service (SaaS), physical resources (Platform as a Service, PaaS), and hardware/infrastructure (Hardware as a Service, HaaS, or Infrastructure as a Service, IaaS) are all examples of cloud computing[4].

A cloud computing service An example is a cloud computing service like Amazon EC2(Amazon Elastic Cloud) [5]uds [6] (Public Clouds, Private Clouds, and Hybrid Clouds) are employed depending on the domain or environment in which they are used (a combination of both private and public clouds) [7].

In Cloud computing, load balancing is critical for an even distribution of nodes in the data center, resulting in improved resource usage and fewer failures. The development of an appropriate load-balancing algorithm is at the heart of these challenges. The load can take several forms, including CPU load, memory capacity, delay, and network load [8].

The technique of dispersing distributed system's load among its nodes to maximize resource efficiency and task response time while avoiding a situation in which some nodes are severely burdened while others are idle or doing very little work is known as load balancing. Load balancing guarantees that all processors in the system or each node in the network are carrying out roughly the same amount of work at any given time. However, problems and roadblocks have arisen in the implementation of load-balancing [9]. The study's main objectives are as follows: Examine the many load balancing methods that have been published in the literature. Evaluate different load balancing techniques, and present an overview of existing load balancing challenges and issues and prospective research opportunities for enhancing load-balancing procedures. User demands change a lot on the cloud necessitating the deployment of a dynamic environment to fulfill the activities.

A cloud service broker determines resource availability and interacts with other brokers related to resource performance and cost when the cloud detects requests from the userbase. After examining the available resources, the broker sends the user request to the designated data center where the Data Center Controller (DCN) accepts it for further processing.

2. CLOUD COMPUTING

Cloud computing (CC) is the supply of computing services over the Internet including servers, databases, storage, networking, and software, cloud computing enables faster innovation, more flexible resources, and lower costs [10]. The customer usually only pays for the cloud services used, which helps save money and runs the infrastructure more efficiently [11], as shown in figure 1.



Fig. 1: Cloud Computing

2.1 Components of Cloud

The following are the three basic components of cloud computing [7], as shown in figure 2:

1. Client Computers: Client computers allow end-users to interact with the cloud.
2. Distributed Servers: The servers are dispersed across the globe, but they work with one another.
3. Data Centers: A data center is a grouping of servers.

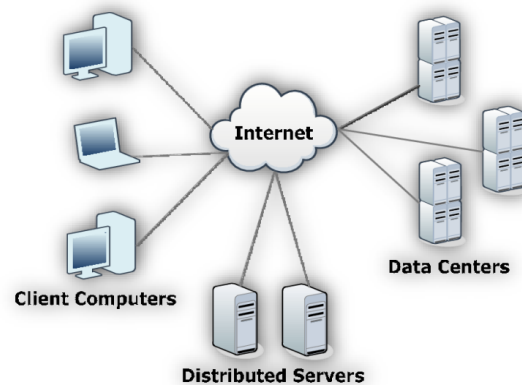


Fig. 2: Cloud Computing Components [12]

3. LOAD BALANCING

In cloud computing, load balancing is the process of redistribution of workload. The basic goal of load balancing is to keep any single server from becoming overburdened and perhaps failing [13][14], In cloud computing, load balancing is an appropriate solution to several issues that develop during the setup and operation of the system. Load balancing in a distributed environment must

consider two main tasks: resource provisioning or resource utilization and task scheduling.

Resources will be immediately available on demand, and resources will be efficiently utilized under high and low load conditions, thanks to efficient resource provisioning and task scheduling. Energy is conserved while the load is low is the technique of distributing workload Equally among virtual computers to maximize Resource proficiency. This study highlights different load-balancing approaches based on different criteria. As a result of the rising demand and load in the field of cloud computing, we saw the need for load balancing among cloud resources. As a result, existing research has been identified and reported in a systematic manner that highlights concerns and challenges for future research efforts based on existing research. A load balancer assists in the equal distribution of resources to activities for optimal resource utilization and user satisfaction at the lowest cost, motivating us to detect load balancing challenges and concerns to resolve them. As shown in figure 3.

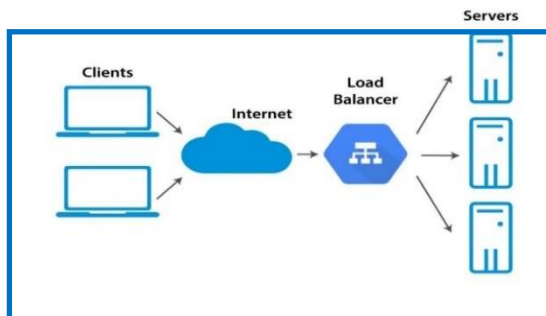


Fig. 3: A load balancing diagram

Load balancing algorithms can be classified into three sorts of methods based on the algorithms: sender-initiated, receiver-initiated, and symmetrically-initiated. The symmetrically-initiated technique combines the two previous methods. Sender initiated: The procedure starts with the sender. The sender sends the request message until finds a receiver who can handle the workload [15]. As shown in figure 4.

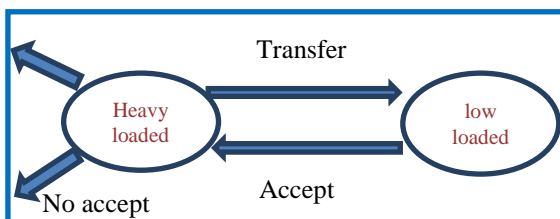


Fig. 4. Sender Initiation

Receiver initiated: The interaction begins with the reception. The receiver sends a message of invitation to a sender who has agreed to share responsibility. As shown in figure 5.

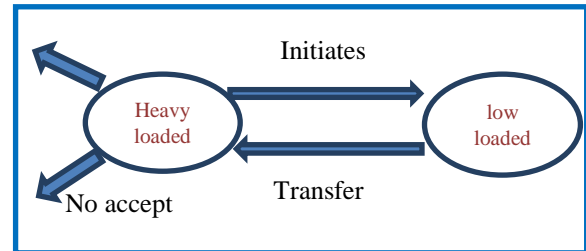


Fig. 5: Receiver Initiation

3.1 Algorithm for Load Balancing

In a Cloud Computing context, load balancing is a critical activity for maximizing resource use. We reviewed numerous load balancing strategies in this study, each with its own set of advantages and disadvantages. On the one hand, static load balancing schemes make simulation and monitoring of the environment easier, but they don't account for the heterogeneous nature of the cloud. Dynamic load balancing algorithms, on the other hand, are harder to simulate but are best suited for cloud computing's heterogeneous environment. This section will go over the different types of load balancing algorithms and how they are compressed [16] as shown in figure 6. Load balancing algorithms can be divided into two groups based on the present condition of the system [17]

3.1.1 Load Balancing in Static Mode

Prior knowledge about the system, such as processing power, memory, execution, and information about the user's requirements is already known in static algorithms. These algorithms do not require any information about the current status of the system. This sort of algorithm has significant drawbacks in the event of a system resource breakdown, and jobs cannot be relocated during execution for load balancing. Round Robin is a static load balancing mechanism that evenly distributes traffic among servers. Several flaws in the round-robin method have been identified, and a new approach called Weighted Round Robin has been developed to address these shortcomings. The basic idea behind this algorithm is that each server is given a weight, and then each server is ranked according to that weight [18].

Advantages of static load balancing algorithms

- Very effective in terms of costs and expenses, as they do not need to monitor the resources (nodes and servers) at the time of execution.
- Most suitable for stable and homogeneous environments as the operating characteristics do not change with time and the payloads are usually identified and fixed.

The disadvantage of static load balancing algorithms

- Inflexible and unable of responding to dynamic changes in the features during execution time.
- It makes allocation decisions without considering the present state of the system.

3.1.2 Dynamic Load Balancing (DLB)

DLB is a term that refers to the process of balancing loads. Load balancing decisions are made by dynamic algorithms based on the current state of the system, and no prior knowledge of the system is required, which mitigates the disadvantages of a static method. Although dynamic algorithms are difficult to implement, they can improve performance and fault tolerance.

3.2 Comparison of Load Balancing Algorithms

Advantages of dynamic load balancing algorithms

- More flexible
- Most suitable for dynamic and heterogeneous environments and can respond to dynamic system changes.

The disadvantage of dynamic load balancing algorithms

- More complex due to the environment's heterogeneity and dynamic nature.
- More cost because it can adapt to the load at the time of implementation depending on the characteristics of the resource at the time of execution.

In the dynamic load balancing algorithm, some policies are employed. These can be summed up as follows: [19] [20]

1. Transfer Policy: A job is chosen for transfer from a close node to a remote node using a dynamic load balancing technique.
2. Processor Selection Policy: the number of processors involved in the load exchange is specified.
3. Location Policy: In a load balancing algorithm, the determination of a destination node for a transfer task is known as location policy or location strategy.
4. Information Policy: In a load balancing algorithm, information policy or information strategy refers to the collection of information about the node in the system. As shown in table 1.

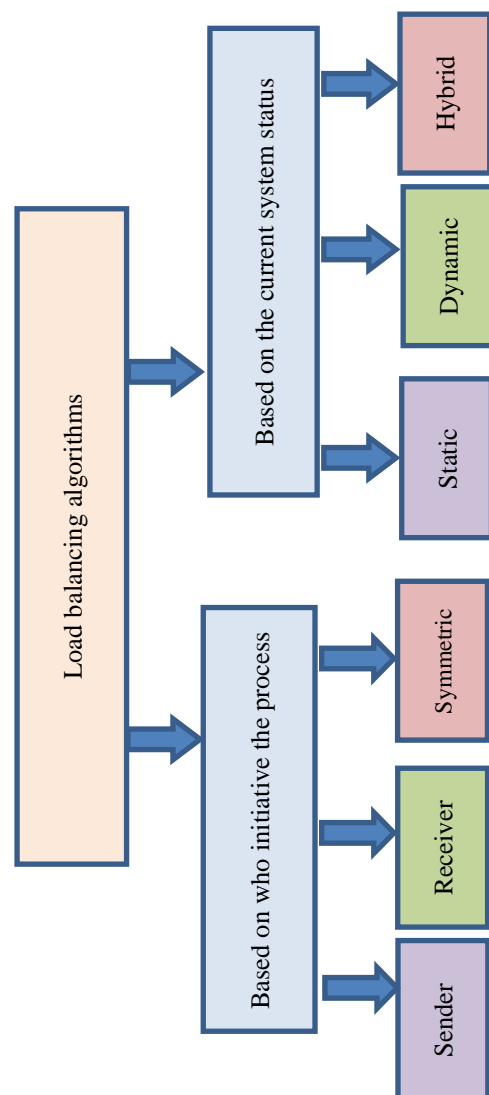


Fig. 6: Classification of load balancing algorithms

Table 1: A comparison of load balancing algorithms

Static load balancing algorithms	Dynamic load balancing algorithms
Used for systems with little incoming load variability.	Used for systems with high incoming load variability.
Communicating with the servers in real-time is not necessary.	It necessitates active real-time contact with the servers.
The traffic is distributed evenly among the servers.	The servers' traffic is dynamically distributed.
During runtime, the allotted load cannot be shifted to other servers.	To reduce resource underutilization, the allotted load might be retransferred among servers.
It needs more detailed knowledge of the existing system resources.	It does not necessarily require prior knowledge of system resources.
Very is effective in terms of costs and expenses, as they do not need to monitor the resources (nodes and servers) at the time of execution this feature overcomes the challenges of load balancing algorithms.	Based on the current state of the system that is used to control the load to overcome the problem of increasing the traffic of the system.
Example: Round Robin, Weighted Round Robin, and Randomized algorithms for load balancing.	Example: Least Connection, Central Queue and Local Queue algorithm for load balancing

3.3 Why Is Load Balancing So Important in A Cloud Environment? [21]

Cloud load balancing is a way of evenly allocating excess dynamic nearby workload among all nodes. Load balancing is used to achieve high user satisfaction and resource utilization ratio by ensuring that no single node is overpowered, hence improving the system's overall performance. When load balancing is utilized correctly, it achieves optimal resource use while reducing resource consumption. Executing failover, enabling scalability, avoiding bottlenecks, avoiding

overprovisioning, minimizing response time, and attaining Green Computing in clouds are all major benefits of employing load balancing. It is achieved through the following factors:

1. Consumption of minimum amount of energy Load balancing reduces energy usage by preventing overheating of equipment due to excessive workload, as several nodes or virtual computers may be lost.
2. Carbon Emissions Reduction Because energy consumption and carbon emissions are exactly proportional, they are inextricably linked. Because load balancing reduces energy usage, it also reduces carbon emissions, thus assisting in the achievement of Green Computing.

3.4 LBA Problems and Challenges in Cloud Computing [22] [23]

The fundamental issue in developing an optimal load balancing technique is that it is extremely difficult to meet all of the users' expectations in terms of improving each load balancing measure. Some users might prefer a faster response time above retaining request execution priority. On the other hand, certain users may demand a tight execution order based on the order in which requests are processed.

Some users want to increase the system's total life span, while others want. It is to be highly available. As a result, to get the best possible outcomes, a compromise between available metrics should be made, and the best, most appropriate load balancing techniques should be developed to enhance reaction time and optimize resource allocation.

In terms of the time it takes to execute and the number of tasks it must complete there are a variety of complex load balancing algorithms. These intricate algorithms result in longer processing times and a more complicated process. Furthermore, implementing such load balancing strategies necessitates extra monitoring, communication overhead, and delays, resulting in more bottlenecks and a reduction in overall system efficiency. Worst-case scenarios should be considered while designing load balancing solutions.

1. The Node's Geographical Distributions

It is used in a variety of largescale programs, like Twitter and Facebook. The DS of the processors is incredibly important in preserving

system effectiveness and ensuring fault tolerance in a cloud computing environment.

2. Algorithm Complexity

The overall performance of a system is influenced by the complexity of load balancing algorithms. Although the method is simple, it has the potential for poor execution in terms of movement time, fault tolerance, and reaction speed which complicates the method. The computations, on the other hand, take longer but yield better outcomes in terms of asset use and throughput.

3. Analysis of Traffic in Different Geographical Locations

The load balancer system maintains activity in each position during peak hours to maximize throughput and asset usage. It is vital for any LBA to assess the activity stream in real-time across several geographic locations and to adapt the workload accordingly.

4. Migration of Virtual Machines

Virtualization allows the entire machine to be seen as a single file or a set of files, which can be used to empty a physical machine that has been overloaded or to swap a virtual machine within a physical machine that has become overloaded. The main purpose is to spread the load across a data center or a group of data centers.

5. Management of Energy

The economy of scale is one of the benefits of energy management that promotes the use of the cloud. Power conservation is the most important feature of a global economy in which the pool of global resources is supplied by a small number of providers rather than everyone having their private services.

6. Single Point of Failure

Non-distributed approaches are used to create different dynamic load balancing algorithms, with load-balancing decisions made by the central node. If the central device fails, the entire computing environment would be affected. As a result, various distributed algorithms that do not rely on a single node to govern the entire computing system must be developed.

7. Heterogeneous Nodes

During early cloud load balancing studies, researchers considered homogeneous nodes. In cloud computing, user requirements vary regularly, forcing execution on heterogeneous nodes to

maximize resource efficiency and reduce reaction time. As a result, designing appropriate load balancing solutions for varied contexts is a challenge for academics.

8. Load-Balancer Scalability

The on-demand availability and scalability of cloud services enable users to access services at any time and scale up or down quickly. A professional load balancer should consider quick changes in demand in terms of processing power, storage, system structure, and so on to facilitate these changes efficiently.

9. Storage Management

Cloud hosting has solved the problem of classical storage systems, which required personal management and had high hardware costs. Users can store a variety of data on the cloud without fear of access issues. Cloud storage is rapidly expanding, requiring the storage of data replication for optimal data access and consistency. Full data replication strategies are wasteful due to the redundant data storage policy on replication points. Partial replication may be sufficient, but dataset availability may be an issue, making load-balancing solutions more complicated. As a result, a load balancing strategy based on a partial replication system that handles the distribution of application and related data must be designed [24].

10. Increased web traffic

The solution to this challenge: The first alternative is a server change, which requires temporarily transferring requests to a more powerful server. The second option is outsourcing, in which queries are forwarded to another server that can respond quickly. The third option is to utilize numerous servers with an effective load balancer that distributes the load evenly among them.

11. job assignment and resource allocation

The solution to these challenges: Provisioning and scheduling of resources are the two fundamental activities of resource management. The user sends a job to the broker with varied QoS limitations during resource provisioning. The broker is in charge of locating an appropriate resource to complete the work. The assignment of tasks and their execution on the most appropriate resource is the essence of resource scheduling underloaded, normal loaded, and

overloaded were the three categories. assigned to the resource. Only underloaded or normally loaded resources are assigned to this job by the scheduler.

12. Load measurement

The solution to these challenges: To manage a load of available resources divided them into three categories: underloaded, normal loaded, and overloaded. That resource provisioning approach can effectively balance resource load. They demonstrated peer load balance, a tool that assesses demand and resource capacity while taking into account the needs of both the client and the service provider. Compared to other current approaches, the provided mechanism lowered the cost and average reaction time. A dynamic threshold value based on standard deviation was established for load balancing and task transfer. The resources are classified for work migration, and the processing element's threshold value is compared to the average load of each cluster. Tasks are chosen at random from underloaded or overloaded resource collections for load balancing.

13. Cost optimization

The solution to this challenge: Adaptive workflow scheduling (AWS) has been proposed, which takes into account the cost of resources as well as the cost of communication between tasks and resources. The final cost is determined regularly due to the varied character of resources, demonstrating a strategy for process activity-based cost optimization. To check the technique efficiency, Variability factor, activity time, and resource usage are used to calculate processing cost and waiting time. A cost optimization strategy for a hybrid cloud was described. Based on resource availability, clouds are divided into three categories: public, private, and hybrid. A user can make use of the public cloud's services on a pay-per-use basis. Individuals own private clouds, which provide a range of free services. In a hybrid cloud, resources from the public cloud are pooled together as needed.

14. Fault tolerance

The solution to these challenges: Resource failures were addressed, and a checkpoint-based task execution recovery method was established. If a task is not completed by the deadline, another acceptable resource is selected to do it. Before being transferred to another suitable resource, the task state is saved and resumed for future execution through the checkpoint. This reduces execution time, reaction time, and throughput when compared

to other existing approaches. included a fault-tolerant scheduling system that guarantees task completion. The replication limitation is overcome by rescheduling the task for later execution. If a failure is detected, the scheduler reassigns the task to another suitable resource, reducing resource waste. This strategy reduces resource consumption and task execution time. uses a fault index to keep track of resource history. The fault rate is calculated based on task completion success and failure on a certain resource. Based on the fault index value, the grid broker repeats the work that can be used when a problem occurs.

3.5 LBA'S Metrics [25]

Load balancing among various resources is a major challenge in cloud computing. By equally allocating jobs to available resources, load balancing assures optimal resource usage. To determine if the load balancing method is the most optimal in terms of resource allocation and task completion time, various measurements must be taken [26].

1. Throughput: This refers to the total number of jobs that have been accomplished over a certain period. For a better system execution, it is necessary to have a high throughput.
2. Associated Overhead: This term refers to the amount of overhead that occurs during the load balancing algorithm's execution. It is made up of task development and it interposes communication. The least amount of overhead is required for the load balancing strategy to perform successfully.
3. Fault-tolerant: This is defined as the capacity to load balance using the proper manner without incurring subjective interface or node failure. A solid approach to fault tolerance should be included in every load balancing algorithm.
4. Migration time: This is the time it takes to move a preparation from one system node to another for execution. This period should be reduced to improve the system's performance.
5. Response time: In a distributed system, response time refers to the time it takes for a given load balancing mechanism to respond. For optimum results, this time should be kept at a minimum.
6. Resource Utilization: This is the parameter that provides information about how the asset is currently being used. The optimal asset should be used for effective load balancing in the system.

7. Scalability: This refers to a load balancing algorithm's ability to handle a system made up of a limited number of processors and machines. For better system performance, this parameter should be enhanced.

8. Performance: This refers to the system's overall efficiency. The overall system performance can be increased if all of the parameters are addressed.

4. CONCLUSION

The purpose of this research was to discuss load balancing problems in a cloud

environment and how to solve this problem. In cloud computing, the LBA is critical for removing overburdened workloads and providing equal and approximate service. In this paper, we discussed many LBA that can be used to improve utilization and better understand the user's needs with many parameters. This paper also discovered that load balancing is a critical issue in cloud systems for balancing storage on-demand services, and data centers, among other things. It is also attained minimum server overhead, maximum resource usage, throughput, reduce processor migration time, improve system performance and efficiency.

REFERENCES

- [1] S. Afzal and G. Kavitha, "Load Balancing in Cloud Computing – A Hierarchical Taxonomical Classification," *Journal of Cloud Computing*, Vol. 8, No.1. Springer, Dec. 01, 2019. doi: 10.1186/s13677-019-0146-7.
- [2] P. Srivastava and R. Khan, "A Review Paper on Cloud Computing," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, Vol. 8, No. 6, p. 17, 2018, doi: 10.23956/ijarcsse.v8i6.711.
- [3] F. Khoda Parast, C. Sindhav, S. Nikam, H. Izadi Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Comput. Secure*, Vol. 114, p. 102580, 2022, doi: <https://doi.org/10.1016/j.cose.2021.10251>.
- [4] Soumya Ray, "Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment," *International Journal on Cloud Computing: Services and Architecture*, Vol. 2, No. 5, pp. 1–13, Oct. 2012, doi: 10.5121/ijccsa.2012.2501.
- [5] T. Alam, "IAIC Transactions on Sustainable Digital Innovation (ITSDI) Cloud Computing and its role in the Information Technology", [Online]. Available: <https://pandawan.aptisi.or.id/index.php/att/article/view/59>, pp. 108-115, No. 2, Vol. 1, 2020.
- [6] W. Ahmad, A. Rasool, A. R. Javed, T. Baker, and Z. Jalil, "Cyber security in IoT-based cloud computing: A comprehensive survey," *Electronics*, Vol. 11, No. 1, p. 16, 2021.
- [7] R. Balasubramanian and M. Ramadhan, "Security Issues: Public vs Private vs Hybrid Cloud Computing General Terms," 2012, Vol. 55, No. 13.
- [8] Y. M. Teo and R. Ayani, "Comparison of Load Balancing Strategies on Cluster-based Web Servers." 2014 International Conference on Computer and Communication Technology (ICCCCT). IEEE, 2014.
- [9] K. Mishra and S. K. Majhi, "A Binary Bird Swarm Optimization Based Load Balancing Algorithm for a Cloud Computing Environment," *Open Comput. Sci.*, Vol. 11, No. 1, pp. 146–160, 2021, doi: 10.1515/comp-2020-0215.
- [10] L. Wang et al., "Cloud Computing: A Prospective Study," in *New Generation Computing*, Apr. 2010, Vol. 28, No. 2, pp. 137–146. doi: 10.1007/s00354-008-0081-5.
- [11] P. Srivastava and R. Khan, "A Review Paper on Cloud Computing," 2018. [Online]. Available: www.ijarcsse.com, Vol.8, No.6, pp. 17-20.
- [12] Z. Chaczko, V. Mahadevan, S. Aslanzadeh, and C. Mcdermid, "Availability and Load Balancing in Cloud Computing." Vol.14, 2011, pp. 134-140.
- [13] R. Rajeshkannan and M. Aramudhan, "Comparative study of load balancing algorithms in the cloud computing environment," *Indian Journal of Science and Technology*, Vol. 9, No. 20, May 2016, doi: 10.17485/ijst/2016/v9i20/85866.
- [14] P. Sarma, C. Kalita, and V. Deka, "A Survey on Load Balancing Algorithms in Cloud Computing," *Int. J. Comput. Sci. Eng.*, Vol. 7, No. 6, pp. 169–176, 2019, doi: 10.26438/ijcse/v7i6.169176.
- [15] N. K. Mishra, "Load Balancing Techniques: Need, Objectives and Major Challenges in Cloud Computing-A Systematic Review," 2015, Vol. 131, No. 18, pp.11-19
- [16] S. K. Mishra, B. Sahoo, and P. P. Parida, "Load Balancing in Cloud Computing: A Big Picture," *Journal of King Saud University – Computer and Information Sciences*, Vol. 32, No. 2. King Saud bin Abdulaziz University, pp. 149–158, Feb. 01, 2020. doi: 10.1016/j.jksuci.2018.01.003.
- [17] E. Jafarnejad Ghomi, A. Masoud Rahmani, and N. Nasih Qader, "Load-Balancing Algorithms in Cloud Computing: A survey," *Journal of Network and Computer Applications*, Vol. 88. Academic Press, pp. 50–71, Jun. 15, 2017. doi: 10.1016/j.jnca.2017.04.007.
- [18] M. Farik, R. Papers, and N. Shah, "Static Load Balancing Algorithms in Cloud Computing: Challenges & Solutions," *International Journal of Scientific & Technology Research*, Vol. 4, No. 10, 2015, [Online]. Available: www.ijstr.org

- [19] P. Kumar and R. Kumar, "Issues and Challenges Of Load Balancing Techniques in Cloud Computing: A Survey," ACM Computing Surveys, Vol. 51, No. 6, Feb. 2019, doi: 10.1145/3281010.
- [20] B. Amin, B. Yagoubi, and Y. Slimani, "Dynamic Load Balancing Strategy for Grid Computing," Pwaset Vol. 13, May 2006 Issn 1307-6884.
- [21] M. Alam and Z. Ahmad Khan, "Issues and Challenges of Load Balancing Algorithm in Cloud Computing Environment," Indian Journal of Science and Technology, Vol. 10, No. 25, pp. 1–12, Jun. 2017, doi: 10.17485/ijst/2017/v10i25/105688.
- [22] N. S. Raghava and D. Singh, "Open Journal of Mobile Computing and Cloud Computing Comparative Study on Load Balancing Techniques in Cloud Computing," 2014.
- [23] J. M. Abdul-Jabbar, "A Simple Analytic Design Procedure for Lattice Wave Digital Filters with Approximate Linear Phase", *Basrah Journal for Engineering Science*, Vol. 11, Issue: 1, pp. 123-134, 2011.
- [24] J. M. Abdul-Jabbar and O. N. Saadi, "Design and Multiplierless Realization of ECG- Based Gaussian Wavelet Filter with Lattice Structures", *AL-Rafidain Engineering Journal*, Vol. 23 Issue: 4, pp. 172-182, 2015.
- [25] F. J. Harris, "Multirate Signal Processing for Communication Systems", USA, ISBN 10: 0131465112, Prentice Hall, 2004.

تحديات تقنيات موازنة الاحمال في البيئة السحابية: مراجعة

توركان احمد خليل

turkan@uomosul.edu.iq

علا هاني سلطان

ola.20enp21@student.uomosul.edu.iq

قسم هندسة الحاسوب ، كلية الهندسة ، جامعة الموصل

الملخص

لتبسيط الضوء على قضايا وتحديات خوارزمية موازنة التحميل (LBA) في سياق الحوسبة السحابية. ناقش مؤلفو هذا العمل موازنة الحمل ومشاكل موازنة الحمل في الحوسبة السحابية وكيفية حلها، والتقنيات الأساسية المستخدمة للحفاظ على الحمل عبر السحابة، بالإضافة إلى مزايا وعيوب هذه الخوارزميات. علاوة على ذلك، تتم مقارنة استراتيجيات موازنة الحمل الثابتة والديناميكية. يجب أن تتجنب طرق موازنة التحميل في السحابة لمراكز البيانات التحميل الزائد / التحميل المنخفض للجهاز الظاهري، وهو ما يمثل تحديًا في الحوسبة السحابية. تساعد آليات موازنة الحمل في إدارة أعباء العمل. الهدف الأساسي من موازنة الحمل بين الأجهزة هو توزيع أعباء العمل وموارد الحوسبة عبر خادم واحد أو أكثر مع تقليل استهلاك الطاقة وتعظيم استخدام الموارد. الحوسبة السحابية هي نوع من الحوسبة التي يتم فيها توفير موارد قابلة للتطوير عند الطلب وتكون افتراضية في كثير من الأحيان، كما هو الحال في خدمة عبر الإنترنت. لمجموعة متنوعة من الأسباب، بما في ذلك توفير التكلفة والسرعة والكفاءة والأداء والأمان، أصبحت الحوسبة السحابية مؤخرًا خيارًا شائعًا للأفراد والمؤسسات. تستخدم الحوسبة السحابية الإنترنت لتقديم مجموعة متنوعة من الخدمات لأجهزة الكمبيوتر والأجهزة الخاصة بالشركة. يمكن أن تتضمن هذه الموارد تخزين البيانات وقواعد البيانات والخوادم والشبكات والبرامج، من بين أدوات وتطبيقات أخرى. نتيجة لذلك، يجب على المطورين والباحثين في هذه الورقة تصميم وإنشاء موازن تحميل مناسب لأنظمة السحابة المتوازنة والموزعة.

الكلمات الدالة:

الحوسبة السحابية؛ مركز البيانات السحابية؛ توزيع الحمل؛ خوارزمية موازنة الحمل؛ تحديات موازنة الحمل.