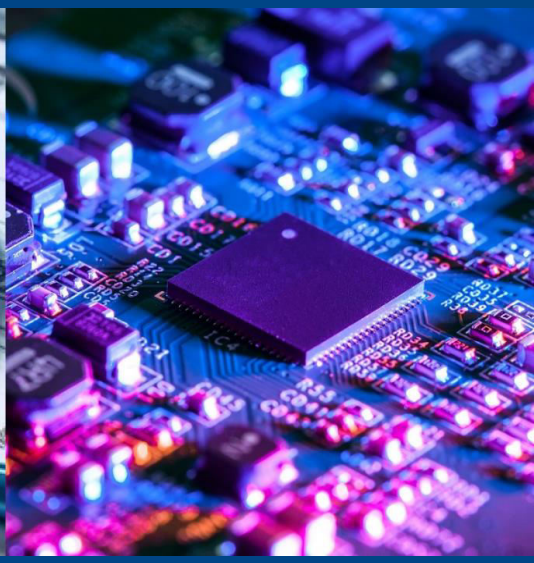
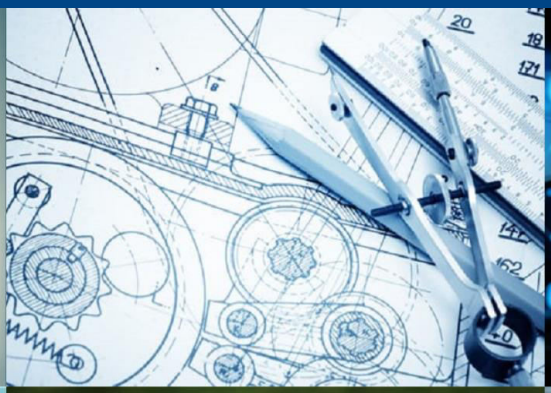
ISSN: 2582-7219

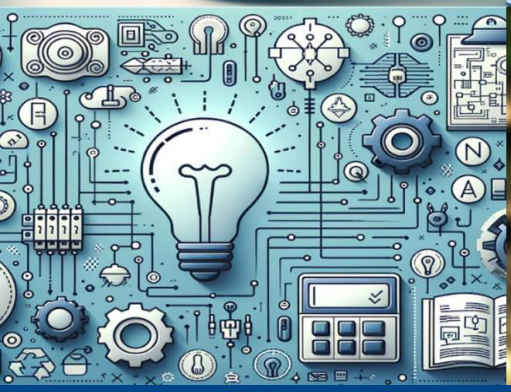




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 4, April 2025**

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Cloud Load Balancing

Pranav Shankar Mane, Prof. Vishwatej M. Pisal

Postgraduate Student, Dept. of Master of Computer Application Anantrao Pawar College of Engineering and Research, Pune, India

Dept. of Master of Computer Application Anantrao Pawar College of Engineering and Research, Pune, India

ABSTRACT: The burgeoning reliance on cloud systems magnifies the necessity for load balancing techniques to adequately distribute ever-swelling workloads. As cloud computing proliferates and customers increasingly insist on scalable, high-performance offerings, research into cloud load balancing has intensified given its strategic role. By judiciously parceling out tasks among numerous nodes, load balancing strives to maximize resource productivity while boosting service quality. But because cloud infrastructures store information in open, distributed systems, the amount of stored data grows exponentially every day, making load balancing all the more vital as a foundational practice for bearing big data's increasingly heavier loads. Proper load management is essential in preventing system overloads and cost ineffectiveness.. Various algorithms have proposed dynamically distributing workloads among cloud nodes, thus stopping resource congestion and boosting system responsiveness. In this scholarly paper, we scrutinize disparate workload allocation methods used in virtual environment and assess their adequacy in job scheduling. A comparative analysis of these techniques offers understandings into recent improvements in this area, emphasizing their potencies and restrictions. Simultaneously, shorter tasks can be allotted to nodes with spare capacity to maximize resource usage while longer computations are distributed more judiciously to prevent nodes from getting overburdened. A well-designed framework must judiciously handle varying task requirements with humming efficiency and also retain flexibility to expand on-demand during high workload periods.

KEYWORDS: Dynamic traffic allocation, Virtual environment, Software-based machine, Resource Optimization, Task Scheduling Algorithms, Cloud Service Scalability

I. INTRODUCTION

Cloud computing delivers on-demand computing capabilities of servers, storage, networking, and software, all via services delivered over the internet. Organizations can access flexible and cost-effective infrastructure without needing to deploy physical hardware on-premises. Since cloud services are offered using a pay-as-you-go service model, users are assured the maximum utilization and operational efficiency of the resources used. Traffic management is the process used in cloud environments to distribute workloads across a number of servers, and cloud environments utilize load balancers for system performance, availability, and reliability. Load balancers sit between clients and backend resources and manage traffic based on the server capability and other conditions at run-time. Load balancers minimize response time by distributing requests across servers with a similar workload, and also prevent potential bottlenecks by maximizing use of resources and improving fault tolerance.

Cloud traffic distribution can be categorized into application traffic distrbution and network traffic distribution. Application load balancing routes requests based upon application layer protocols (HTTP/HTTPS), while network load balancing distributes traffic at the transport layer (TCP/UDP). Large cloud vendors(like Google Cloud Load Balancer, AWS Elastic Load Balancer, and Azure Load Balancer) leverage advanced algorithms to dynamically load balance workloads forbest system performance.Google Cloud Load Balancing takes advantage of Maglev, Andromeda, and Google Front Ends (GFEs) to intelligently distribute traffic to multiple backend instances across the world. Similarly, Microsoft Azure has a range of load balancing products including global and regional traffic distribution systems to operate multi-region workload without sacrificing security compliance.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. PRINCIPLES OF CLOUD LOAD BALANCING

2.1 Balancing Loads:

Cloud traffic refers to the distribution of network traffic among numerous servers or application instances to improve performance, availability, or to avoid a single server being overwhelmed or not serving requests adequately. Rather than directing requests from users to a single server—incurring possible bottlenecks—a load balancer can distribute requests among many Servers.

The main use of load balancing is to maximize use of resources, stability, and ease of scaling. Load balancing in cloud computing plays a key part in achieving continuous availability and error resistance, so that cloud-based systems can non-clinically scale to meet varying workloads. It helps minimize downtime, avoids performance degradation, and increases resource utilization especially when there is peak traffic.

2.2 Significance of Workload Distribution in Cloud Platforms

Traffic Management is a critical factor in the performance and consistency of distributed applications.

The following illustrates the main benefits of load balancing:

1.Improved Performance: Load balancing distributes workloads relatively equally to all servers in a server pool which lowers the amount of processing load on each server thereby improving response times, and efficiency of the system.

2.Reliability and Fault Tolerance: Applying traffic management strategies removes a single point of failure, and in the case of server failure, traffic can be redirected seamlessly. This allows for guaranteed continuity of service.

3. System adaptability: Smart and responsive resource allocation can facilitate scaling up or down, thereby providing systems the ability to handle variability in traffic and consequently a smooth user experience.

4.Efficient Resource Utilization: By being efficient with resource utilization, we can limit the over-provisioning of resources and lessening spending, while creating the most optimal efficiency**.**

2.3 Types of Traffic Distribution

2.3.1 Fixed workload distribution

This type of load balancing evenly distributes traffic on the basis of predefined rules and prior knowledge of what each individual server can do. Algorithms assign a predefined workload to a few designated servers, they work well when the environment is predictable. They do not consider real-time changes in server load, however, and are weak on traffic patterns with high variance.

2.3.2 Traffic distribution in real time

This kind of load balancing bases its decisions about traffic distribution on the network's current state, server performance, and load. Because it is a more adaptive technique and guarantees a balanced workload, it works better in cloud environments with fluctuating demand**.**

2.4 Load Balancing Mechanisms

2.4.1 Application-based traffic distributors

This are applications that run on standard servers, balancing traffic without specialized hardware..

Key Characteristics:

1.Uses existing hardware, thus requires lower initial outlay.
2.Can add more virtual instances as needed without hassle.
3.You can adjust the settings of the application according to its requirements.
4.Integrates directly and easily into a cloud system.

Use Cases:

1.Ideal for dynamic and cloud-based applications.
2.Ideal for places where configuration changes often.
3.Cost-effective for small to mid-scale deployments.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.4.2 Hardware Load Balancers.
Hardware load balancers are Purpose-specific devices used to handle network traffic at high speed.

Key Characteristics:
1.Special hardware made for load balancing, high performing
2.Can handle large volumes of traffic while causing less latency
3.Includes Firewall Functions and SSL offloading.
4.It can easily be deployed / configured in the network.

Use Cases.
1.Best suited for large-scale enterprise applications.
2.Used in circumstances which require low latency and high reliability
3.Ideal for associations with strict security needs.

2.5 Design Approaches for Load Management
Load distribution is a critical concern for resource utilization and system performance in modern distributed systems. The two main architectures for load management are Centralized and Decentralized. The sections below offer an extensive overview of these two different approaches, discussing their features, benefits and challenges.

2.5.1Centralized Load Balancer Systems :
Centralized load balancing system employs a hierarchical module in which one or only a few of the central servers are used to manage the dispersion of workloads through the whole structure. This process enables a better-organized and a more regulated atmosphere for traffic distribution and resource allotment.

Key Characteristics:
1.Centralized Control: At a high-level, all processing, scheduling, and distribution of traffic is centrally controlled.
2.Less Complexity in Management : Easier to set up, configure and monitor due to a single decision-making unit.
3.**Efficient Resource Allocation:** The centralized system can easily allocate resources via defined rules making it efficient in many aspects from performance to security

Challenges:
1.Scalability Problems: With increasing demand, the central server might eventually become a bottleneck and thus limits the system scalability.
2.Central Server Down: If the central server fails, it can knock the entire system down for the largest amount of time. It is primarily used in typical enterprise networks, small data centers, and applications where strict and orderly traffic management is desired.

2.5.2 Decentralized Systems
In decentralized load balancing, workloads are distributed among multiple nodes and has no single point of control. It allows to balance the load among nodes, improving scalability and fault tolerance.

Key Characteristics:
1.Error tolerance: The system continues to function when nodes are present.
2.Scalability: Insert with new nodes dynamically without large reconfiguration
3.Higher Resilience: Adaptive workload distribution reduces over-reliance on a single component.

Use Cases:
1.Applications and services that necessitate continuous uptime and are cloud-native.
2.Dynamic workloads in large-scale distributed environments Reliant on redundancy failover systems

2.5.3 Decentralized Load Balancing Systems
Decentralized load balancing systems disperse the responsibility of monitoring workload management to many nodes across a system, removing the dependence on the success of one point of control. Each node has management responsibility for balancing the load as well maintaining fault tolerance and increasing system scalability.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Key Features:

1.Fault-Tolerant: The system continues to operate at a high level of availability despite node failures.

2.Scalable: Nodes can dynamically integrated into the system based on capacity or threshold without greatly impacting the schema.

3.Adaptive Load Distribution: Load balancing is distributed dynamically at the node level without preordained structure allowing for enough flexibility.

Visual Representation of Load Balancing Architectures:

The following diagram illustrates the fundamental differences between **centralized** and **decentralized** load balancing approaches:

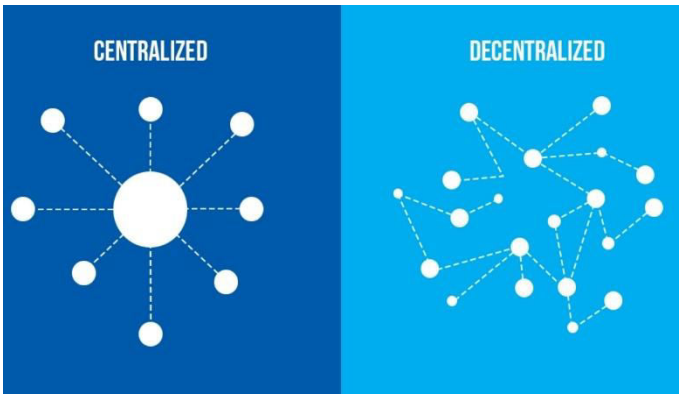


Fig: Comparison of Centralized and Decentralized Load Balancing Architectures

This visual representation clearly differentiates the centralized system, where all connections lead to a single control node, from the decentralized system, where multiple nodes distribute workloads independently.

III. RELATED WORK

Task Distribution Strategies in Cloud Computing

3.1Fixed Load Distribution Algorithms

Fixed load distribution algorithms employs an allocation of jobs according to pre-defined determinants and does not react to inherent changes in the server workload or state. The implementation of these algorithms is straightforward but may give rise to inefficiency when workloads are imbalanced.

3.1.1 Round Robin Method

This algorithm distributes load evenly across servers. Each incoming request is sequentially assigned or handed off to a single server in the pool in turn, without considering the server's current load.

Applications:

1.Ideal for systems where the processing capacity is similar across all servers.

2.Ideal for equally shared workloads and state is not significant, such as web applications with light requests.

Advantages:

1. Easy to implement

2.Simple to manage fair workload is distributed fairly on average to the servers.

Disadvantages:

1.The server workload is not considered.

2.Unbalanced server loads; some servers may be overloaded.

3.The Round Robin method is inefficient when server capacity is heterogeneous.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.1.2 Load Distribution with Weights

By the addition of introducing weights to the servies according to their processing power, the Weighted Round Robin is an extension of the Round Robin method. The servers that have a higher weight/share will proportionally receive a higher number of requests.

Use Cases:

1.Better suited for an environment that is heterogeneous in server resources.
2.Better in a system where specific nodes are able to handle more requests than a some,
3.Pros: Takes into account capacity, lessens the likelihood that low-capacity servers can be overloaded. Provides more flexibility in the traffic distribution adjustments.
4.Cons: Requires constant evaluation and adjustments of weight whenever the server capacity fluctuates.

Advantages:

1.Performance Sensitivity: This method allocates jobs taking into account the ability of each server using weighted allocation.
2.Elastic Behavior: It supports workload variation without a loss in performance seamlessly.

Disadvantages:

1.Increased Intricacy: This method uses a more complicated process than the simple Round Robin method.
2.System Tuning: It requires reconfiguring for weight allocations whenever server performance changes.

3.1.3 IP Hash Method

The principle of the IP Hash algorithm is to apply a hash function to the IP address of the client in order to map their requests to an appropriate server. If this algorithm is applied consistently, each time a client sends requests, they will reach the same server, which provides some persistence. It is ideal for maintaining session persistence within stateful applications. It can work in scenarios that require the client to remain on the same server.

Advantages:

1.The user session is preserved.
2.The mapping of each request to one specific server can reduce repeated computations.

Disadvantages :

1.Redistributing traffic in case of a server failure may result in some inconsistency.
2.There is some added computation time that is required due to the hashing function.

3.2 Dynamic Task Allocation Strategies

This algorithm strategy quickly react to changes based on the system's condition and distribute traffic according to the present condition of the servers. Although these dynamic algorithms make it work better and resource utilization, they come at a cost of additional computation**.**

3.2.1. Least Connection Method

The least connection algorithm sends requests to the server with the smallest number of live connections. This balances the traffic between the servers as requests are distributed.

Applications:

1.Good for applications with long held (or persistent) connections such as video streaming.
2.Works well in environments with variability or fluctuations in requests.

Advantages:

1.Prevents overload of servers with high requests. Can dynamically respond to changes in traffic patterns.
2.Must monitor active connections continuously.
3.More expensive next to static algorithms**.**

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.2.2 Weighted Minimum Connection Strategy

The Weighted Minimum Connection (WLC) approach is a load-balancing strategy that improves upon the Least Connection method by considering the capacity of each server along with the number of active connections.

Applications:

1.For applications that have a long connection time for requests, such as video streaming.
2.For environments that have fluctuating volumes of requests.

Advantages:

1.Helps mitigate overload on heavily strained servers.
2.Dynamically manages factors to changes in traffic patterns.
3.more computationally intensive than static algorithms..

3.2.3 Minimum Response Time Method

The algorithm selects the server with the fastest response time to handle incoming requests and response time, considering both processing speed and active connections.

Applications:

1.Suitable for real-time applications requiring low latency.
2.Useful in cloud environments optimizing for speed and responsiveness.

Advantages:

1.Enhances user experience by reducing response times.
2.Dynamically adapts to fluctuating loads.

Disadvantages:

1.Requires continuous performance monitoring.
2.Can lead to frequent reassignments, adding computational overhead.

3.2.4 Resource-Based Method

The Resource-Based approach allocates the traffic according to server resources that it has available at the moment, such as CPU, memory, bandwidth, etc. Specialized monitoring agents track system measurements, which helps to determine traffic allocation.

Applications:

1.Best suited for cloud environments with mixed workloads.
2.Applied to distributed systems at scale where it is critical to optimize resources.

Advantages:

1.Maximizes the utilization of resources.
2.Optimizes the control of the platform by dynamically adjusting to the actual server load.

Disadvantages:

1.Have to have sophisticated monitoring tools.
2.Computational overhead from communicating resource information for updates to real-time status.

IV. METHODOLOGY

4.1Challenges of Load Distribution

It is an important area of cloud technology that allows workloads to be balanced across distributed computing resources. That said, there are several obstacles encountered when trying to implement load balancing that must be considered. Through research and academic investigation, the following are challenges regarding load balancing:

1.Geographically Distributed Nodes:

Cloud computing infrastructure relies on data centers or servers spread across the world. For these distributed nodes are working together as a system to facilitate user requests, these nodes must act as a coordinated, unitary system. There

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

are existing methods of load balancing that mainly focus on a local region and disregard factors like communication delays, network latencies, or available user or resource locations within the request. The difficulty increases further when nodes are in increasingly remote locations or accessibility to traditional algorithms simply won't work. It is important to have advanced algorithms to manage geographically distributed nodes while keeping the latency and minimizing congestion.

2.Non-uniform System

In a hosted environment, user demands vary significantly, necessitating utilization of diverse computing nodes. Load-balancing algorithms historically focused on homogeneous environments, leading to inefficiencies in handling varied computational capabilities. The lack of effective algorithms for heterogeneous systems presents a significant challenge. Further research is required to develop dynamic load-balancing strategies that accommodate diverse processing power, memory capacity, and resource availability.

3.Virtual Machine Migration

Virtualization facilitates the operation of numerous virtual machines (VMs) on a singular physical machine for dynamic resource allocation. However, when a physical machine reaches its capacity, VMs must migrate over to another server machine using live migration techniques. This process requires a lot of bandwidth, thus moving a large VM image over a network with limited bandwidth would take a long time. The implementation of filtering and data reduction techniques would alleviate this concern by optimizing host workload distribution. Moreover, allocating additional bandwidth for the purpose of VM migration can assist in overcoming network bottleneck that would assist in seamless migration.

4.Scalability of Load Balancers

Scalability by definition means that users can scale their resource usage up or down, on-demand as they see fit in cloud computing. A suitable load balancer must efficiently manage variation in compute demand (CPU), storage demand (I/O capacity), and capacity of the cluster as a whole. Implementing load-balancing algorithms that scale the system as required to meet dynamic workload requirements is necessary to maintain the optimal performance of the system.

5.Complexity of Load-Balancing Algorithms

The efficiency of a cloud system is heavily influenced by the complexity of its load-balancing algorithms. While some methods appear simple in design, they may perform poorly in terms of migration time, fault tolerance, and response speed. Load-balancing techniques should be continuously refined to optimize resource distribution and system performance, regardless of workload variations.

6. Storage Management

Traditional storage systems require costly hardware and extensive management, but cloud computing leverages distributed storage across multiple nodes. Effective data replication is crucial to maintaining accessibility and redundancy. While full replication can be inefficient due to high storage costs, partial replication introduces challenges in load balancing and dataset availability. Developing advanced load-balancing methods tailored to partial replication systems can enhance data distribution and application performance.

7.Single Point of Failure

Several existing load balancing algorithms confront a single machine to provide a responsibility of managing load balancing among the multiple VMs. This exposes the system to a possible point of failure, as the failure of a single machine can render the entire system unusable. To address this, distributed load balancing algorithms need to be designed to provide fault tolerance and resiliency.

8.Security and Quality of Service (QoS)

Cloud environments need to enforce high levels of security and Quality of Service (QoS) standards. Load-balancing algorithms should have security techniques embedded to mitigate unauthorized access, identify anomalies, and distribute resources equitably. Researchers are looking for ways to improve load balancing while not compromising data integrity, VM security, or service quality.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. RESULT AND DISCUSSION

5.1. Market Overview

Market Overview According to Future Market Insights, the cloud load balancer market was valued at USD 7.76 billion in 2022 and is projected to grow to USD 33.08 billion in 2033, expanding at a CAGR of 15.6% during the forecast period of 2023-2033. This impressive growth can be primarily attributed to three function factors: the extensive adoption of cloud computing solutions, the surging demand for security and scalability, and cost effectiveness..

5.2. Technological Advancements Driving Market Growth

The cloud load balancing market is being driven by technology. Cloud computing provides businesses with the ability to utilize cloud computing for advanced analytics and improved data management. When confronted with limited server capacity or performances, automated load balancing solutions are taking the place of manual load balancers to provide better scalability with far less administrative workload. A good part of this is helpful for e-commerce business particularly when traffic spikes happen during peak season shopping like Black Friday. Compared to traditional load balancers, cloud software is typically more cost effective for flexibility options, and even has additional cloud-based options, such as autoscaling features, while keeping applications online. Additionally, cloud-based load balancers have great configuration options to work with other cloud services giving an overall better user experience with total network management..

5.3 Key Industry Players and Market Trends

Major Market Participants and Industry trends multiple major technology companies are making investments in cloud load balancing solutions. Market leaders are as follows:
Amazon Web Services (AWS),Microsoft Google, Cloud, IBMF5, Networks Citrix Systems, NGINX,Radware,Fortinet,Kemp Technologies For instance, Google Cloud Load Balancing offers powerful network load balancing and traffic management functions for applications run on Google Compute Engine and Google Kubernetes Engine. Red Hat and Microsoft doubled down on their relationship by integrating OpenShift with Microsoft Azure, yielding a more effective cloud infrastructure.

5.4. Market Dynamics

1.Drivers of Market Growth Increased Awareness and Adoption: The cloud networking is being adopted by enterprises operating in the government, defence, and retail sectors for centralized management and security.
2.Increasing Online Traffic: As global internet usage climbs, load balancing has become more efficient.
3.Cloud Networking Trends: Organizations transaction their traditional networking models to cloud-based architectures for more operational efficiency.
4.Demand for AI-Based Automation: AI and machine learning are being integrated into network management, facilitating real-time traffic distribution and security monitoring.

5.5. Future outlook

the most promising technology advancements in hosted environment is cloud load balancing. As businesses of all sizes become more reliant on cloud solutions, an increasing number of companies, large and small, will be using cloud-based load balancers because of the cost savings, scalability, and reliability they promise. Also, investments in AI-powered load balancing, along with network automation, will continue to accelerate growth in this market. As businesses like Hetzner are rolling out their automated load-balancing solutions to maximize efficiency and distribute traffic loads effectively. Online global traffic will continue to increase, and so will the need for virtual environment solutions. Overall, Virtual environment will be very significant in the cloud-computing future..

VI. CONCLUSION AND FUTURE WORK

6.1.Summary of Key Findings

Conclusion and Future Work Summary of Main Findings For the rationale of network traffic optimization, performance improvement, and scalability, Cloud traffic management is necessary in cloud environments. The study noted that cloud traffic management can enhance application reliability, security, and cost savings. The large cloud service providers (e.g. Amazon, Google, and Microsoft) will continue to develop new automated and AI technologies for cloud.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804543

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

6.2 Limitations of the Study

This study was based on secondary data and is therefore not necessarily a la carte to alth distances contemporaneously. Cloud load balancing can be impacted by factors such as integration complexities, security issues, and network latency consistent with cloud load balancing efficiencies.

6.3 Recommendations for Future Research

Future studies should focus on new AI technologies to improve load balancing. Additionally, studies could look to assess load balancing principles following factors including edge computing or 5G, or provide empirical case studies of various cloud (and application) load balancing approaches. Cloud load balancing will continue as an important technology and any recent developments or future enhancements will improve efficiencies, reliability and scalability on an ongoing process.

REFERENCES

1.Google Cloud. (n.d.). Cloud Load Balancing Overview. Retrieved from https://cloud.google.com/load-balancing/docs/load-balancing-overview

2. DeJonghe, Derek. **Load Balancing in the Cloud**. O'Reilly Media, 2018

3.Journal of cloud computing: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-019-0146-7

4.ACM Digital Library

5. zenarmor website: https://www.zenarmor.com/

6. Future Market Insights: https://www.futuremarketinsights.com/

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com