

Cloud Data Centre Balancing Load by using Cloud Service Algorithm

PROJECT SYNOPSIS ON

Cloud Data Centre Balancing Load by using Cloud Service Algorithm

BACHELOR OF TECHNOLOGY

INFORMATION TECHNOLOGY



Roll No.-2001320130043

Name-Jyoti Gupta

Branch -IT

4th year

SUPERVISOR/GUIDE

Deepak Mangal

Ms. Uma Tomer

Greater Noida Institute of Technology, Greater Noida

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

October,2023

INDEX

(i). List of Figure.....	2
1.Abstract.....	1
2.Introduction.....	2
2.1System Requirement.....	1
2.2 Module.....	1
3.Literature Review.....	2
4.Objective.....	2
5.Methodology.....	2
6.Conclusion.....	1
7.References.....	1

ABSTRACT

Despite the many past research conducted in the Cloud Computing field, some challenges still exist related to workload balancing in cloud-based applications and specifically in the Infrastructure as service (IaaS) cloud model. Efficient allocation of tasks is a crucial process in cloud computing due to the restricted number of resources/virtual machines. IaaS is one of the models of this technology that handles the backend where servers, data centers, and virtual machines are managed. Cloud Service Providers should ensure high service delivery performance in such models, avoiding situations such as hosts being overloaded or underloaded as this will result in higher execution time or machine failure, etc. Cloud service algorithm contributes to load balancing, and scheduling tasks much adheres to the requirements of the Service Level Agreement (SLA), a document offered by cloud developers to users. Important SLA parameters such as Deadline are addressed in the LB algorithm. The proposed algorithm is aimed to optimize resources and improve Load Balancing in view of the Quality of Service (QoS) task parameters, the priority of VMs, and resource allocation. The proposed LB algorithm addresses the stated issues and the current research gap based on the literature's findings. Results showed that the proposed LB algorithm results in an average of 78% resource utilization compared to the existing Dynamic LBA algorithm. It also achieves good performance in terms of less Execution time .

Introduction

Cloud data centers are critical infrastructures that provide various computing services to users over the internet. As the demand for cloud services continues to grow, data centers face the challenge of efficiently balancing the load across their resources to ensure optimal performance and resource utilization. To tackle this challenge, cloud service algorithms have been developed to dynamically distribute workloads and optimize resource allocation within the data center.

Cloud service algorithms employ intelligent load balancing techniques to ensure that incoming requests are efficiently distributed among available resources. These algorithms consider factors such as resource capacity, current utilization levels, and network conditions to make informed decisions about workload distribution. The primary goal is to achieve high performance, minimize response time, and avoid resource overutilization or underutilization.

Some commonly used cloud service algorithms for load balancing in data centers:

Round Robin: This algorithm distributes incoming requests sequentially among available resources. Each resource is allocated the same number of requests in a circular manner, ensuring a fair distribution. It is a simple approach but does not consider the current workload or resource utilization, which may lead to imbalanced loads.

Weighted Round Robin: This algorithm extends the Round Robin approach by assigning weights to each resource based on its capacity or performance capabilities. Resources with higher weights receive a proportionally larger share of the workload, enabling better load balancing based on their capabilities.

Least Connection: This algorithm directs incoming requests to the resource with the fewest active connections at a given time. It aims to evenly distribute the workload based on the current resource utilization, ensuring that no resource becomes overwhelmed with requests.

Random: The Random algorithm assigns incoming requests to available resources randomly. While simple to implement, it may not provide effective load balancing, as it does not consider resource capacities or current utilization levels.

Response Time-based: This algorithm measures the response time of each resource and assigns incoming requests to the resource with the lowest response time. By prioritizing the resources with faster response times, it aims to optimize user experience and reduce overall latency.

Application-aware: This algorithm considers application-specific factors, such as the type of workload or the specific requirements of the application, when distributing requests. It allows for more intelligent load balancing decisions based on the characteristics of the workload being processed.

These are just a few examples of cloud service algorithms used for load balancing in data centers. In practice, data center operators may employ a combination of these algorithms or develop custom algorithms tailored to their specific requirements. The choice of algorithm depends on the data center's workload characteristics, resource capacities, network conditions, and performance objectives.

By leveraging these cloud service algorithms, data centers can effectively balance the load across their resources, ensuring optimal performance, scalability, and resource utilization, thereby providing efficient and reliable cloud services to users.

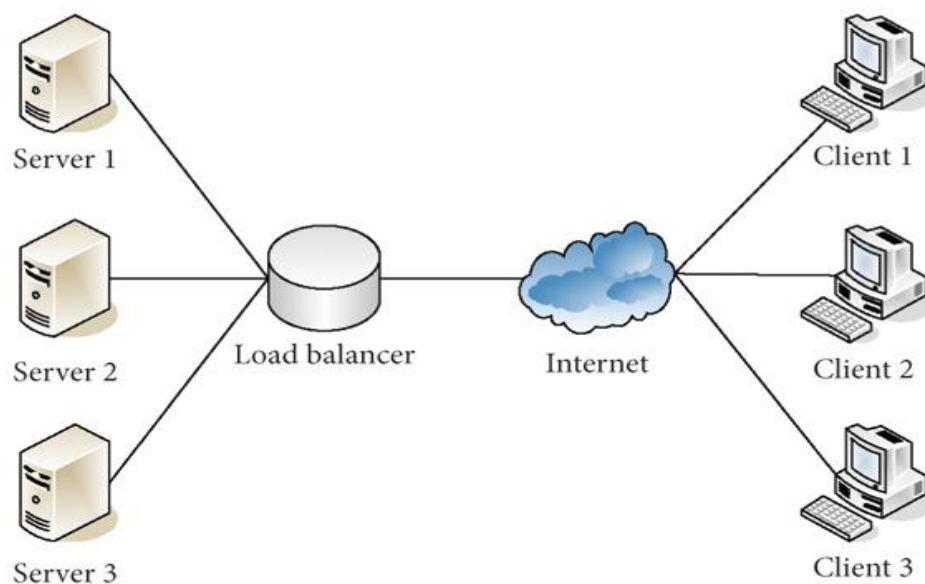


Fig1: Load Balancer Architecture

2.1 System Requirement

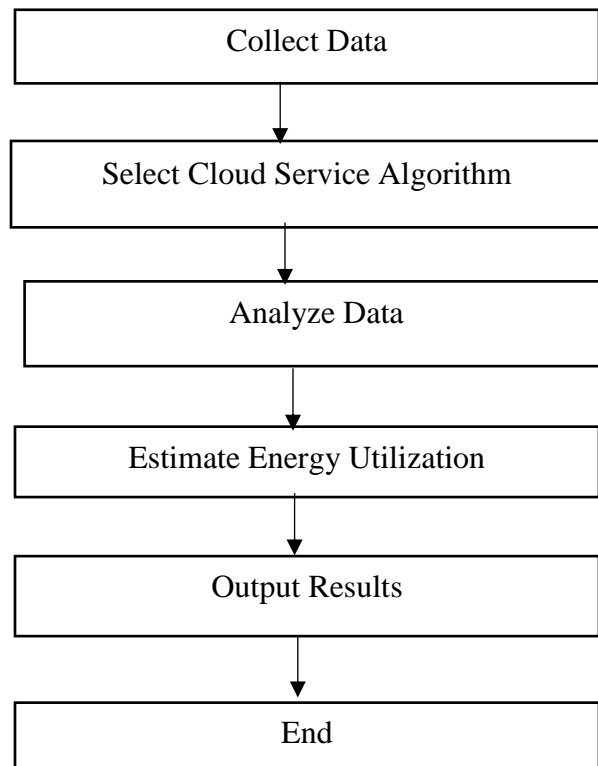
Software Requirement

- Energy Monitoring Software
- Database Management System
- Data Analytics Tools
- Resource Allocation Software
- Resource Balancing Algorithms
- Efficiency Metrics Calculation

Hardware Requirement

- Energy Meters and Sensors
- High Speed Network Equipment
- High-Performance CPUs
- Memory :- 2GB

2.2MODULE



Literature Review

1.Title: "Cloud Computing Load Balancing Mechanism Taking into Account Load Balancing Ant Colony Optimization Algorithm" [2]

Authors: Jing He

Published: 2022

Load balancing is realized by monitoring the usage of resources on each physical node and then analyzing the collected resources; by analyzing the algorithm of load balancing, the load on the node is dynamically migrated to the node with a small load; and finally the purpose of load balancing is achieved. Any computer or server source can be optimized, the response time can be as short as possible, and the processor performance can be as high as possible. Under normal circumstances, the user partition is finally obtained by analyzing the source of each controller (such as CPU usage, memory space, and bandwidth).

2.Title: "A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications" [1]

Authors: Dalia Abdulkareem Shafiq, Noor Zaman Jhanjhi; Azween Abdullah,

Mohammed A. Alzain

Published: 2021

This Paper explains the proposed and improvised Load Balancing in Cloud Computing Environment. This algorithm's primary goal is to provide services of high quality to clients in Cloud Computing applications. The method consists of both processes: Task Scheduling process to assign deadline and completion time to cloudlets (tasks) and secondly, Load Balancing process to perform migration of workload in case of VM violation to maintain a balanced load in the cloud environment.

3.Title: "Designing a Component-Based Throttled Load Balancing Algorithm for Cloud Data Centers" [3]

Authors: Dawit Mekonnen, Alemayehu Megersa ,Rakesh Kumar Sharma and Durga Prasad Sharma

Published: 2022

In this research paper, efforts have been made to answer the three important questions: What are the deficiencies in the existing load balancing algorithms used for load balancing over cloud data centers? Which parameters can be explored to enhance the performance of the throttled algorithm?

How to design an enhanced component-based throttled algorithm for improved response time, processing time, and resource utilization with a comparison matrix?

4.Title: "Dynamic Load Balancing Algorithm for Cloud Data Centers Based on Virtual Machine Migration" [5]

Authors: F. Liu, H. Wu, and J. Li

Published: 2017

This research paper proposes a dynamic load balancing algorithm for cloud data centers that utilizes virtual machine migration. The algorithm considers the CPU and memory utilization of virtual machines and migrates them to less loaded hosts, thereby achieving load balancing. The paper presents simulation results showing improved performance in terms of resource utilization and response time.

5.Title: "Dynamic Load Balancing Strategy for Cloud Data Center based on Multi-objective Optimization"

Authors: W. Xu, J. Zhang, and B. Li

Published: 2016

This research paper proposes a dynamic load balancing strategy for cloud data centers based on multi-objective optimization. The algorithm considers factors such as resource utilization, response time, and energy consumption. It employs a genetic algorithm to find an optimized load balancing solution. The paper presents experimental results demonstrating improved load balancing efficiency compared to traditional algorithms.

OBJECTIVE

The objective of balancing the load in a cloud data center using cloud service algorithms is to achieve efficient resource utilization, optimal performance, and high availability. The specific objectives can include:

1.Load Distribution: The primary objective is to evenly distribute the workload across available resources, ensuring that no resource is overutilized or underutilized. This helps prevent bottlenecks and maximizes resource utilization, leading to efficient operation of the data center.

2.Performance Optimization: Load balancing algorithms aim to optimize the performance of cloud services by minimizing response time, reducing latency, and improving throughput. By intelligently distributing workloads, these algorithms ensure that resources are efficiently utilized, resulting in improved user experience and satisfaction.

3.Scalability: Load balancing algorithms should be designed to handle increasing workloads and scale the data center resources accordingly. This objective involves dynamically allocating resources based on the current demand and workload characteristics, allowing the data center to scale up or down seamlessly.

4.Fault Tolerance: Load balancing algorithms should consider fault tolerance to ensure the high availability of cloud services. By monitoring the health and availability of resources, these algorithms can detect failures and redirect workloads to alternative resources to minimize service disruption.

5.Cost Optimization: Efficient load balancing helps optimize resource usage, which can lead to cost savings. By avoiding resource overprovisioning or underutilization, cloud data centers can optimize their infrastructure costs and achieve better return on investment.

6.Energy Efficiency: Load balancing algorithms can contribute to energy efficiency in data centers by distributing workloads to minimize the overall power consumption. By consolidating workloads onto a reduced number of active resources and powering down underutilized resources, energy savings can be achieved.

7.Adaptability and Dynamic Decision Making: Load balancing algorithms should be adaptable to changing conditions, such as variations in workload patterns or resource availability. They should be capable of making dynamic decisions in real-time to adjust the workload distribution based on current resource utilization, performance metrics, and network conditions.

In summary, the objective of load balancing in cloud data centers using cloud service algorithms is to achieve efficient resource utilization, optimal performance, high availability, scalability, fault tolerance, cost optimization, energy efficiency, and adaptability. These objectives collectively contribute to providing reliable and efficient cloud services to users while maximizing the benefits of cloud computing infrastructure.

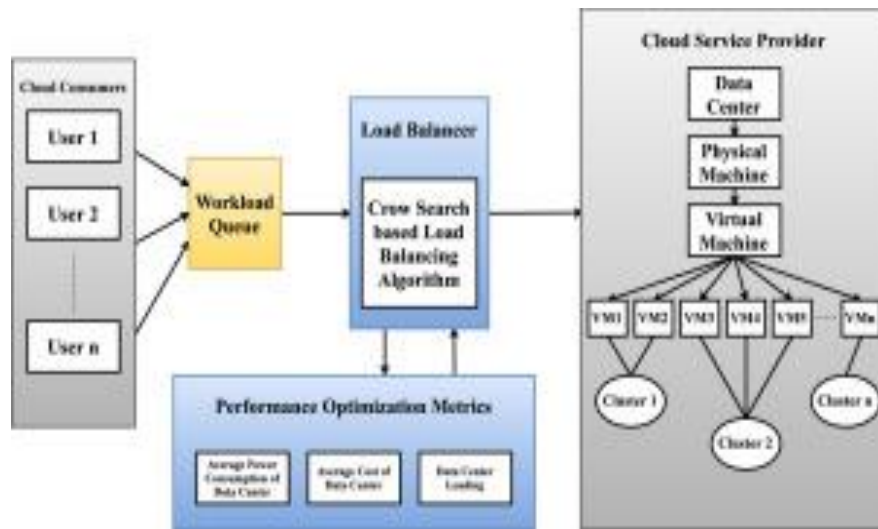


Fig 2: Load Balancing working

METHODOLOGY

The methodology for balancing the load in a cloud data center using cloud service algorithms typically involves the following steps:

1.Data Collection and Monitoring: Gather relevant data about the current state of the data center, including resource utilization metrics (e.g., CPU, memory, network), workload characteristics (e.g., incoming requests, traffic patterns), and performance metrics (e.g., response time, latency). This data can be collected through monitoring tools, sensors, or performance measurement systems.

2.Algorithm Selection: Choose the appropriate load balancing algorithm based on the specific requirements, objectives, and constraints of the data center. Consider factors such as workload characteristics, resource capacities, network conditions, performance goals, and scalability needs. Different algorithms may be suitable for different scenarios, so select the algorithm that best fits the data center's requirements.

3.Load Distribution: Implement the selected load balancing algorithm to distribute incoming workloads across available resources. This involves making decisions about where to route each incoming request or task based on the algorithm's logic. The algorithm may consider factors such as resource availability, current utilization levels, response time, and workload characteristics.

4.Resource Allocation and Scaling: Adjust resource allocation based on the load balancing decisions. This may involve provisioning additional resources (e.g., virtual machines, containers) to handle increased workloads or dynamically scaling down resources during periods of low utilization. Resource allocation can be automated using orchestration tools or cloud management platforms.

5.Performance Monitoring and Feedback: Continuously monitor the performance of the data center and gather feedback on the effectiveness of the load balancing algorithm. Measure key performance indicators (KPIs) such as response time, throughput, resource utilization, and user satisfaction. This feedback loop helps evaluate the algorithm's performance and identify areas for improvement.

6.Algorithm Optimization and Adaptation: Analyze the collected data and feedback to optimize and adapt the load balancing algorithm. This may involve fine-tuning algorithm parameters, incorporating machine learning techniques to predict workload patterns, or considering additional factors for decision-making (e.g., energy efficiency). Regularly evaluate and refine the algorithm to enhance its performance and align with changing workload characteristics or resource availability.

7.Testing and Evaluation: Conduct thorough testing and evaluation of the load balancing algorithm in a controlled environment or using simulations. Assess its effectiveness in achieving the desired objectives such as load distribution, performance optimization, scalability, fault tolerance, and cost efficiency. Compare the algorithm's performance against other algorithms or benchmarks to identify its strengths and limitations.

8.Deployment and Continuous Monitoring: Deploy the load balancing algorithm in the production environment of the cloud data center. Continuously monitor the performance and stability of the algorithm, ensuring that it effectively balances the load, adapts to changing conditions, and meets the data center's objectives. Regularly collect and analyze data to detect any anomalies or performance degradation, and make necessary adjustments to maintain optimal load balancing.

By following this methodology, cloud data centers can effectively balance the load across their resources and optimize the performance of their cloud services.

CONCLUSION

Balancing load in a cloud data center using cloud service algorithms is a fundamental practice for maintaining high performance, availability, and scalability. By intelligently distributing workloads, these algorithms contribute to the efficient use of resources, cost savings, and improved user experiences. However, it's crucial to choose the right algorithm for specific use cases and continuously monitor and adjust the load balancing strategy to adapt to changing demands and challenges in a cloud environment.

In summary, load balancing using cloud service algorithms is an indispensable component of modern cloud computing, enabling businesses to deliver reliable and responsive services to their users while optimizing resource utilization and maintaining security.

REFERENCES

1. Dalia Abdulkareem Shafiq, Noor Zaman Jhanjhi; Azween Abdullah, Mohammed A. Alzain (2021). A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications.
2. Jing He (2022). Cloud Computing Load Balancing Mechanism Taking into Account Load Balancing Ant Colony Optimization Algorithm.
3. Dawit Mekonnen, Alemayehu Megersa ,Rakesh Kumar Sharma and Durga Prasad Sharma (2022). Designing a Component-Based Throttled Load Balancing Algorithm for Cloud Data Centers”
4. Choudhury, S., Misra, S., Roy, S., & Mukherjee, A. (2015). A Survey on Load Balancing Algorithms for Cloud Data Centers. *Journal of Network and Computer Applications*, 45, 1-14.
5. Liu, F., Wu, H., & Li, J. (2017). Dynamic Load Balancing Algorithm for Cloud Data Centers Based on Virtual Machine Migration. *Journal of Computational Science*, 20, 172-182.
6. Buyya, R., Ranjan, R., & Calheiros, R. N. (2009). *CloudSim: A Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services*. Elsevier, 41(1), 23-50.
7. Dua, S., Goswami, R. K., & Kumar, P. (2014). A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment. *International Journal of Computer Applications*, 85(3), 1-6.