Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
June 2024 5(6), 1139–1145

PERFORMANCE ANALYSIS & OPTIMIZING CLOUD STORAGE USING A DYNAMIC WORKLOAD ASSESSMENT

Safeena Ansari[1], S. Veenadhari[2]

[1] Ph. D. Scholar, Computer Application, Rabindranath Tagore University, Bhopal (M.P.), India
[2] Professor, Computer Science & Engineering, Rabindranath Tagore University, Bhopal (M.P.), India

Corresponding Author
Safeena Ansari,
safeena_au222284@aisectuniversity.ac.in

DOI
10.29121/shodhkosh.v5.i6.2024.4568

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Cloud storage has become a fundamental component of modern computing, offering scalable and cost-effective solutions for data management. However, optimizing cloud storage performance while handling dynamic workloads remains a significant challenge. This paper explores Dynamic Workload Assessment and Performance Analysis as a strategy to enhance cloud storage efficiency. We analyze workload variations, including read/write operations, latency, and storage utilization patterns, to develop adaptive optimization techniques. Machine learning algorithms and predictive analytics are leveraged to anticipate workload fluctuations and allocate resources dynamically. Additionally, we evaluate various storage optimization strategies such as caching, duplication, compression, and tiered storage management to enhance performance and reduce costs. Experimental results demonstrate that dynamic workload-aware optimizations significantly improve cloud storage responsiveness, throughput, and resource utilization. The study concludes with key recommendations for designing intelligent, self-optimizing cloud storage systems that ensure scalability, efficiency, and cost-effectiveness in dynamic computing environments.

Keywords: Dynamic Load Assignment, Cloud System, Load Optimization, Data Center Load

1. INTRODUCTION

Improving storage performance is essential in today's cloud computing environment to guarantee smooth operations and satisfy user demands. "Dynamic Workload Assessment & Performance Analysis of Cloud Storage Optimization" offers a calculated method for improving storage performance by using dynamic workload assessment methods.

Cloud storage services have become essential for both consumers and businesses in the age of exponential digital growth and worldwide data generation. Large volumes of data can be stored in scalable, adaptable, and affordable ways with cloud storage. But in the midst of so many services, evaluating and comprehending the performance traits of these platforms is essential for making well-informed decisions and making the most use of available resources. The two components of cloud storage optimization—dynamic workload assessment and performance analysis—are closely related, and both greatly enhance the general efficacy and efficiency of storage systems. Here's how they enhance one another.

Real-time monitoring and analysis of workload trends and demands are part of dynamic workload assessment. We use this data to dynamically distribute resources, such as processor power, storage capacity, and network bandwidth, based on variations in workload. For instance, we can allocate additional resources to important tasks to ensure peak performance during periods of high demand.

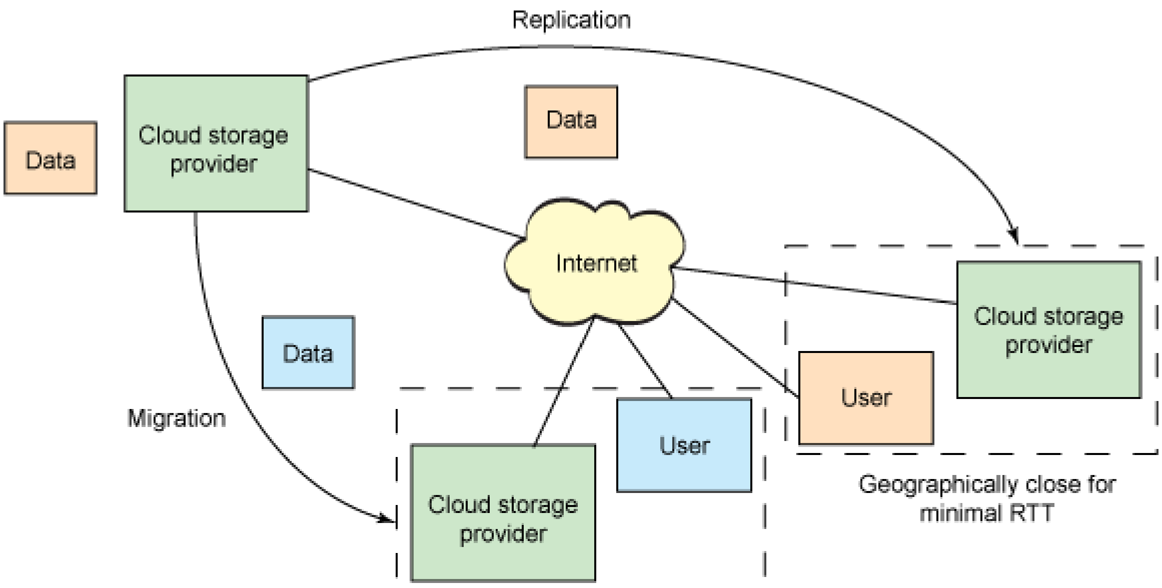

Figure 1: Scalability of cloud storage

It is important to be able to scale both the storage itself (functionality scaling) and the speed to that storage (load scaling). Another important thing about cloud storage is that it can spread data across different locations, so users can access it from the data center that is closest to them. We refer to this as "geographic scalability." Similar to content delivery networks, it is also possible to copy and distribute data that is only readable. Figure 4 shows this.

2. RELATED WORK

This section examines the contributions of authors in the domains of load optimization and load balancing in cloud computing. All research efforts are efficacious and contribute to equilibrating the user request load in the cloud.

This research [1] presents "A Multi-Objective Based Scheduling Framework for Effective Resource Utilization in Cloud Computing" introduces a viable methodology for optimizing resource allocation in cloud systems. The framework employs multi objective optimization methods to tackle the intricate trade-offs associated with maximizing resource utilization, minimizing latency, and decreasing energy usage. The framework's design and decision-making algorithms exhibit adaptability to fluctuating workloads and scalability for various cloud applications.

The study [2] "A Comparative Analysis of Meta heuristic Techniques for High Availability Systems" came out. It is current and useful. This study looks into how well meta heuristic techniques work at making sure that systems are highly available, which is an important part of current computer infrastructure. Comparing and analyzing different meta heuristic algorithms helps the study demonstrate their pros and cons when it comes to system reliability and fault tolerance.

The article [3] "Server placement in mobile cloud computing: a comprehensive survey for edge computing, fog computing, and cloudlet" looks in depth at how to place servers in mobile clouds using edge computing, fog computing, and cloudlets, among other paradigms. The poll looks at many things, including how to best use resources, cut down on latency, save energy, and make things bigger or smaller. It shows the difficulties and chances of putting servers in different places to make mobile cloud services run faster and more reliably.

The article [4] Performance, scalability, availability, and security in different cloud environments are looked at. The article shows a complicated picture where trade-offs and strategic factors are critical. When it comes to scalable resource sharing and high availability, public clouds are the best. However, shared infrastructure security may make people nervous. Private clouds offer enhanced control over speed and security, yet they require meticulous management and incur significant costs. Hybrid clouds take the best parts of both public and private clouds and combine them. They improve speed and availability while lowering security risks by spreading out work in a smart way.

The article [5] "Cloud service selection as a fuzzy multi-criteria problem" It's hard to make a choice when choosing cloud services because performance, cost, security, and compliance are all fuzzy multi-criteria problems that all need to be solved. The uncertainty and personal tastes that come up when choosing a cloud service are hard for traditional decision models to handle. As a way to make decisions that is similar to how people make decisions, fuzzy logic is a hopeful idea.

Researchers usually test how well, reliably, and scalable cloud-API-based apps are when they connect to cloud services through APIs as part of an experimental performance review [6]. This test measures things like reaction times, throughput, resource use, and error rates under various loads and conditions. Researchers can look into how well cloud-API-based apps work by running controlled tests and gathering real-world data.

The assessment of server less applications and infrastructures entails evaluating critical metrics such as response time, scalability, resource utilization, and cost efficiency. Server less computing provides benefits such as automatic scaling, less operational expenses, and pay-as-you-go pricing structures [8]. Performance may fluctuate due to factors such as function complexity, concurrency, and cold start latency.

Modelling caching techniques to improve data access and storage efficiency in cloud environments is part of simulating and evaluating cloud storage caching for science that uses a lot of data [9]. To study caching methods, cache hit rates, data transfer speeds, and resource use under different workloads, researchers use simulation tools. They can find out how better data retrieval works, how much it costs, and how scalable data-heavy scientific applications can be by simulating real-life situations and testing caching strategies like Least Recently Used (LRU) or Least Frequently Used (LFU).

A performance test of cloud-API-based apps includes a lot of testing and analysis to see how well, reliably, and scalable they are when using APIs to connect to cloud services [10]. To find things like reaction times, throughput, resource utilization, and error rates, researchers use different workloads and situations in controlled experiments. By gathering real-world data and comparing the outcomes, they find performance bottlenecks, find the best ways to use resources, and provide well-informed suggestions for improving the system's general performance and the user experience.

The overview and empirical study on file correlation in cloud storage look at how files are saved, accessed, and linked in cloud storage systems as a whole [11]. This study looks into things like data duplicates, file similarity analysis, and how to store things efficiently in the cloud. Collecting real-world data and using statistical analysis to learn about file correlation patterns, storage optimization techniques, and how these things affect the performance and cost of storage are all part of empirical studies. By studying how file correlation changes over time, researchers can come up with better ways to manage storage, speed up data retrieval, and make the best use of resources in cloud storage infrastructures. This will make cloud-based storage options more scalable and cost-effective in the long run.

An RSA-based secure cloud storage protocol using OpenStack is being tested to see how well it works, how safe it is, and how well it can be scaled up in a cloud computer setting [12]. As part of this review, things like encryption and decryption speeds, data transfer rates, system overhead, and resource use with different loads and conditions are being measured. Researchers can study the protocol's performance, find possible bottlenecks or vulnerabilities, and improve system configurations to improve security and performance by running tests and gathering real-world data in an OpenStack environment.

A major part of measuring the performance of cloud service providers and the cloud supply chain is looking at key metrics like service availability, reliability, responsiveness, scalability, and cost-effectiveness [13]. This evaluation includes the monitoring and analysis of the performance of individual cloud service providers, as well as the overall dynamics of the supply chain, which involves data transmission speeds, interactions between providers, and service level agreements (SLAs).

A cloud database security algorithm's success is judged by how well it keeps sensitive data safe, reduces security holes, and makes sure that security rules are followed in a cloud setting [14]. As part of this review, things like encryption and decryption speeds, access control, data integrity checks, and the amount of work that the system has to do to run the database are measured.

In cloud computing, the quantitative analysis and performance review of target-orientated replication strategies include a close look at how data replication strategies affect performance metrics like resource use, data availability, and reliability [15]. This review looks at various replication methods, such as local replication, remote replication, and hybrid approaches, to see how well they meet specific goals, like reducing access latency, increasing data durability, or making the best use of resources.

The comprehensive evaluation of the capabilities and limitations of various cloud environment systems is necessary to evaluate their performance, scalability, availability, and security [16]. To guarantee optimal system responsiveness, performance evaluation encompasses the measurement of factors such as response times, throughput, and resource utilization under various duties. Scalability analysis is concerned with the capacity of cloud systems to accommodate growing demands by dynamically scaling resources while maintaining performance levels.

The fuzzy multi-criteria problem of choosing cloud services means figuring out how to pick the best cloud services based on a lot of different criteria that may not all be clear or important at the same time [17]. During this decision-making process, factors such as performance, cost, scalability, security, compliance, and vendor reputation must be considered. Additionally, subjective preferences and imprecise data must be taken into account.

This research [18] presents a load balancing and energy optimization technique (user-to- server) in cloud computing (LEOCC). The user submits a request to the core switch, which initially assesses the availability of the service by evaluating the load on each core switch. If the service is unavailable, the aggregator switch receives the request. They assess the load of each aggregator switch and choose the least burdened switch for the service request. In the event of service unavailability, we escalate the request to the subsequent level, the access switch.

3. PROPOSED METHODOLOGY

Analyzing incoming workload dynamically involves continuously monitoring and interpreting various metrics related to resource usage, access patterns, and performance indicators in real-time or near-real-time. Here's some algorithms and techniques that can be used for dynamic workload analysis in cloud storage environments:

3.1 THRESHOLD-BASED MONITORING:

Thresholds for key performance metrics can be used such as CPU utilization, disk I/O, and network bandwidth. Continuously monitor these metrics, triggering alerts or actions when thresholds are exceeded. For example, if disk I/O exceeds a certain threshold, it may indicate heavy read/write operations, prompting the system to allocate more resources or adjust caching strategies.

3.2 ANOMALY DETECTION:

Machine learning algorithms can be used to detect anomalies in workload behavior. Train models on historical workload data to establish normal patterns. Continuously compare real-time data with the learned patterns to identify deviations that may indicate abnormal workload behavior or performance issues. Anomalies could signify potential resource contention, security threats, or other issues that require attention.

3.3 PATTERN RECOGNITION:

Pattern recognition techniques can be used to identify recurring workload patterns and their impact on system performance. Classify workloads based on access patterns, request types, or user behavior. Tailor optimization strategies and resource allocations to specific workload patterns, such as prioritizing caching for frequently accessed data or optimizing data placement.

STEP TO DYNAMIC WORKLOAD ASSESSMENT

An approach to explore "Dynamic Workload Assessment for Cloud Storage Performance Optimization" could involve:

1. Selection of Cloud Storage Providers: Choose a few well-known cloud storage providers (like AWS S3, Google Cloud Storage, and Azure Blob Storage).
2. Basic Metrics: Focus on basic performance metrics such as upload/download speed, latency, and availability.
3. Testing Methodology: Design simple tests like uploading and downloading sample files of different sizes to measure the time taken and latency.
4. Comparative Analysis: Compare the results obtained from different providers and analyses their performance based on the metrics.
5. Cost Consideration: Consider basic cost factors associated with storage, egress charges, and any other relevant pricing metrics.
6. Documentation and Presentation: Compile all findings in a straightforward report or presentation format, highlighting the observed differences and any notable insights.

4. RESULT ANALYSIS

The management of server traffic is a challenging undertaking in the presence of users who make frequent requests. This part compares the results of the new dynamic workload distribution of cloud storage (DWCS) method to those of the Heros, Green, and Round Robin methods. The proposed DWCS is more efficient and performs better in managing network traffic.

4.1 SERVER LOAD HANDLING ANALYSIS

When operating in a cloud environment, it is essential to effectively manage server load in order to guarantee performance, availability, and cost efficiency. When it comes to auto scaling, the process of automatically adding or removing instances depending on real-time demand is not necessarily an easy operation. For the purpose of avoiding overload, incoming traffic should be distributed over numerous instances. Optimize queries and indexing to limit the amount of data that is stored in the database for proper load handling. In the figure 2 the load handling capability of DWCS are better as compared to rest of all approaches. The load handling percentage is 93% which is 13% more from round robin and 4% more from recent Heros approach. The better load handling means better server utilization.

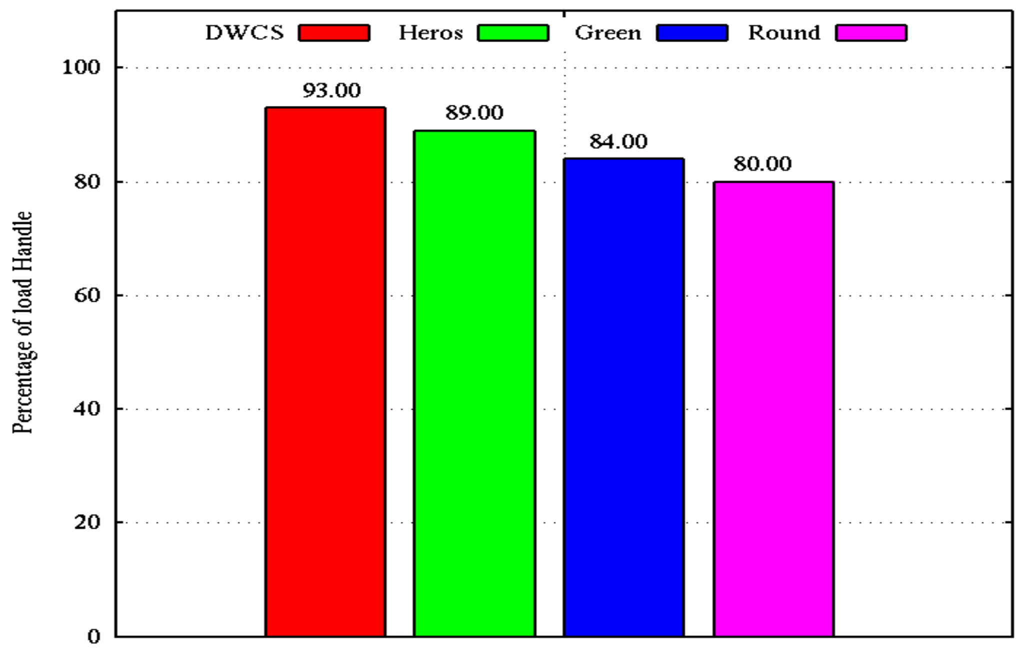

Figure 2: Load Handling Analysis of Servers

4.2 CLOUD NETWORK OVERHEAD

When discussing cloud computing, the term "overhead" refers to the additional expenses, resources, or inefficiencies that are incurred as a result of the management and operation of cloud-based technologies. It is possible for this overhead to be administrative, computational, operational, or financial in nature. The utilization of resources that is not optimum results in increased costs. Virtual machines, storage, or databases that are not being exploited to their full potential are known as idle resources. In the figure 3, proposed DWCS are showing only 0.12 overhead in cloud network and rest of the approaches overhead performance is poor. For thr proper utilization of resources proper balancing is the key factor. To eliminate unnecessary overhead and leverage the benefits of the cloud by optimizing their resources, employing automation, and adopting managed services.

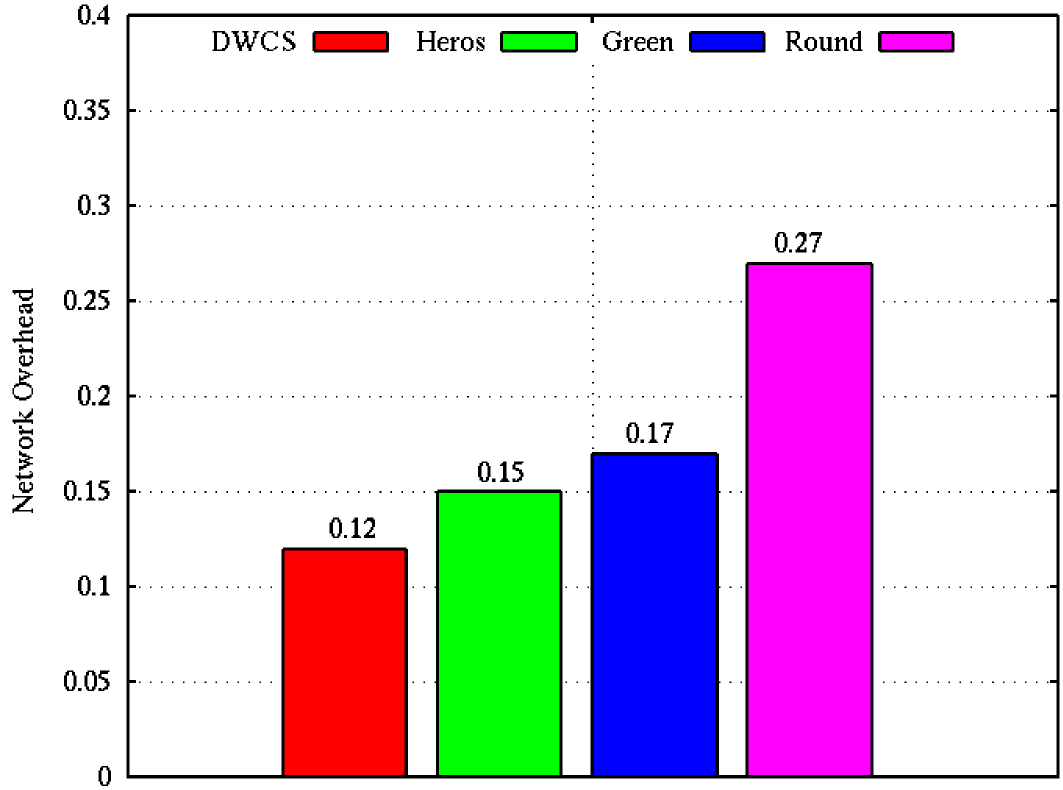

Figure 3: Network Overhead Analysis

5. CONCLUSION & FUTURE WORK:

This research aims to contribute valuable insights and recommendations for users, enterprises, and decision-makers. The results will guide in selecting the most suitable cloud storage service aligned with specific performance requirements, thereby optimizing resource utilization and enhancing operational efficiency. To evaluate the performance of the proposed DWCS algorithm in terms of data center load and network overhead, we compare it to other scheduling methods, including Round-Robin, Green, and Heros. The proposed DWCS works well on the cloud network in terms of throughput, energy use, load balancing, better storage efficiency, less delay, and handling loads.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Reddy, Pillareddy Vamsheedhar, and Karri Ganesh Reddy. "A Multi-Objective Based Scheduling Framework for Effective Resource Utilization in Cloud Computing." IEEE Access (2023).

Syed, Darakhshan, et al. "A Comparative Analysis of Metaheuristic Techniques for High Availability Systems (September 2023)." IEEE Access (2024).

Asghari, Ali, and Mohammad Karim Sohrabi. "Server placement in mobile cloud computing: a comprehensive survey for edge computing, fog computing and cloudlet." Computer Science Review 51 (2024): 100616.

Kumar, Athmakuri Naveen. "Analysis On Performance Scalability Availability And Security In Various Cloud Environment Systems." The journal of contemporary issues in business and government 27.1 (2021): 4583-4591.

Ilieva, Galina, et al. "Cloud service selection as a fuzzy multi-criteria problem." TEM Journal 9.2 (2020): 484.

Alzboon, Ghufran, and Amro Al-Said Ahmad. "A Performance Evaluation Approach for n-tier Cloud-Based Software Services." Proceedings of the 2022 6th International Conference on Cloud and Big Data Computing. 2022.

Ujjainkar, Chirag. "COMPARISION BETWEEN GOOGLE DRIVE, ONE DRIVE AND DROPBOX.,2022"

Scheuner, Joel. Performance Evaluation of Serverless Applications and Infrastructures. Chalmers Tekniska Hogskola (Sweden), 2022.

Wegner, Tobias, et al. "Simulation and evaluation of cloud storage caching for data intensive science." Computing and Software for big Science 6.1 (2022): 5.

Abuzrieq, Yara, Amro Al-Said Ahmad, and Maram Bani Younes. "An Experimental Performance Evaluation of Cloud-API-Based Applications." Future Internet 13.12 (2021): 314.

Ma, Jun. "Overview and Empirical Research on File Correlation in Cloud Storage." Procedia Computer Science 188 (2021): 33-39.

Hyder, Muhammad Faraz, and Syeda Tooba. "Performance Evaluation of RSA-based Secure Cloud Storage Protocol using OpenStack." Engineering, Technology & Applied Science Research 11.4 (2021): 7321-7325.

Azadi, Majid. Performance Measurement of Cloud Service Suppliers and Cloud Supply Chain. Diss. 2021.

Banothu, Srinu, A. Govardhan, and Karnam Madhavi. "Performance evaluation of cloud database security algorithms." E3S Web of Conferences. Vol. 309. EDP Sciences, 2021.

Waseem, Quadri, et al. "Quantitative analysis and performance evaluation of target-oriented replication strategies in cloud computing." Electronics 10.6 (2021): 672.

Kumar, Athmakuri Naveen. "Analysis On Performance Scalability Availability And Security In Various Cloud Environment Systems." Journal of Contemporary Issues in Business and Government Vol 27.01 (2021).

Ilieva, Galina, et al. "Cloud service selection as a fuzzy multi-criteria problem." TEM Journal 9.2 (2020): 484.

K. Chandravanshi, G. Soni, D.K Mishra, "A Method for Load Balancing and Energy Optimization in Cloud Computing Virtual Machine Scheduling," Advances in Intelligent Systems and Computing, vol 1453, pp. 325-335, 2024.