International Research Journal of Engineering Science, Technology and Innovation Vol. 9(3) pp. 1-3, June, 2023
Available online http://www.interesjournals.org/IRJESTI

Review Article

Enhancing Performance and Resource Allocation in Cloud Computing: A Comprehensive Review

Johnny Neck*

Department of computer science, Australia

*Corresponding Author's E-mail: neck_jo8@gmail.com

Received: 01-June-2023; Manuscript No: irjesti-23-103591; **Editor assigned:** 03-June-2023; Pre-QC No: irjesti-23-103591 (PQ); **Reviewed:** 19-June-2023; QC No: irjesti-23-103591; **Revised:** 23-June-2023; Manuscript No: irjesti-23-103591 (R); **Published:** 30-June-2023, DOI: 10.14303/2315-5663.2023.111

Abstract

Cloud computing has emerged as a dominant paradigm in the field of information technology, enabling organizations to efficiently store, process, and manage vast amounts of data. With the rapid growth of cloud services and applications, there is a need to continuously improve the performance and resource allocation mechanisms to ensure optimal utilization of cloud resources. This research article provides a comprehensive review of recent advancements and approaches in enhancing performance and resource allocation in cloud computing environments.

Keywords: Cloud computing, Performance enhancement, Resource allocation, Workload scheduling, Task scheduling, Load balancing, Virtual machine placement, Resource provisioning, Dynamic resource provisioning, Auto-scaling

INTRODUCTION

Cloud computing has revolutionized the way organizations store, process, and manage their data and applications. By providing on-demand access to a shared pool of computing resources, cloud computing offers unprecedented scalability, cost-efficiency, and flexibility (Rosenthal A, 2009). As the demand for cloud services continues to grow rapidly, there is a critical need to continually enhance the performance and optimize the allocation of resources in cloud computing environments. Efficiently utilizing cloud resources is crucial for maximizing performance, ensuring high availability, and meeting the ever-increasing demands of users (Chenna R, 2003). Furthermore, effective resource allocation mechanisms play a pivotal role in maintaining service quality, minimizing response time, and optimizing resource utilization (Altschul SF, 1990). The objective of this research article is to provide a comprehensive review of recent advancements and approaches in enhancing performance and resource allocation in cloud computing. By examining the state-of-the-art techniques and emerging trends, this article aims to shed light on the challenges, opportunities, and future directions in this field.

The article is structured as follows

First, we delve into various performance enhancement techniques employed in cloud computing. This section explores workload scheduling, task scheduling, load balancing, virtual machine placement, and resource provisioning techniques (Wall DP, 2011). Through a comparative analysis of these techniques, their advantages, limitations, and performance implications are discussed. Next, we focus on resource allocation mechanisms in cloud computing. This section reviews dynamic resource provisioning, auto-scaling, and quality of service (QoS) optimization techniques. We analyze their effectiveness in optimizing resource allocation, scaling resources based on demand, and ensuring desired levels of service quality (Wall DP, 2007). Furthermore, we examine the role of machine learning and artificial intelligence in resource allocation and performance optimization. Machine learning algorithms, reinforcement learning, and deep learning techniques have shown promise in making intelligent resource allocation decisions. We discuss their applications, case studies, and experimental evaluations, highlighting their potential to enhance performance and resource utilization in cloud computing environments (Altschul SF, 1990). Security

and privacy considerations are of paramount importance in cloud computing. Thus, we address the challenges and recent advancements in ensuring data security, access control, and privacy preservation. We explore topics such as data encryption, authentication mechanisms, and secure data sharing, which are essential for maintaining trust and protecting sensitive information in the cloud (Abdullah A, 2015). Looking towards the future, we discuss emerging trends and technologies that are shaping the landscape of cloud computing. Edge computing, server less computing, containerization, and hybrid cloud models have gained significant attention (Jouini M, 2015). We explore their potential implications for performance enhancement and resource allocation and highlight the research challenges that need to be addressed.

Performance enhancement techniques

Performance enhancement techniques in cloud computing encompass a range of approaches and strategies aimed at improving the overall performance and efficiency of cloud computing systems. These techniques target various aspects such as workload scheduling, task scheduling, load balancing, virtual machine placement, and resource provisioning (Zhao S, 2013). By optimizing these processes, performance bottlenecks can be mitigated, resource utilization can be optimized, and user experience can be enhanced.

- Workload scheduling: Workload scheduling involves determining the assignment and execution order of tasks or workloads on available resources in the cloud. Techniques such as job prioritization, task clustering and deadline-aware scheduling algorithms are employed to optimize resource allocation and minimize task completion time. By effectively scheduling workloads, resource utilization can be improved, and overall system performance can be enhanced.

- Task scheduling: Task scheduling focuses on the efficient allocation and execution of individual tasks within a workload (Banu SS, 2018). Various scheduling algorithms, such as round-robin, shortest job first, and fair share, are utilized to allocate tasks to available computing resources. By considering factors like task dependencies, resource availability, and task priorities, task scheduling techniques aim to minimize execution time and maximize resource utilization.

- Load balancing: Load balancing ensures an equitable distribution of computational load across available resources in the cloud. Load balancing algorithms and policies distribute incoming workloads among servers or virtual machines to avoid resource overloading and prevent bottlenecks. Dynamic load balancing techniques continuously monitor the system's resource utilization and dynamically adjust the workload distribution to maintain optimal performance.

Resource allocation mechanisms

Efficient resource allocation is crucial for maximizing the utilization of cloud resources while meeting user demands. This section reviews the state-of-the-art resource allocation mechanisms, including dynamic resource provisioning, auto-scaling, and quality of service (QoS) optimization. The challenges and opportunities in resource allocation are explored, along with discussions on energy-efficient resource management.

Machine learning and artificial intelligence in resource allocation

Machine learning and artificial intelligence techniques have shown great potential in improving resource allocation decisions in cloud computing. This section examines the applications of machine learning algorithms, reinforcement learning, and deep learning in resource allocation and performance optimization. Case studies and experimental evaluations are presented to highlight their effectiveness.

Security and privacy considerations

Cloud computing raises concerns regarding the security and privacy of user data and resources. This section addresses the challenges and recent advancements in ensuring data security, access control, and privacy preservation in cloud environments. Topics such as data encryption, authentication mechanisms, and secure data sharing are discussed.

Future directions and challenges

This section discusses the emerging trends, future directions, and open research challenges in enhancing performance and resource allocation in cloud computing. It highlights areas such as edge computing, serverless computing, containerization, and hybrid cloud models, which have the potential to shape the future of cloud computing.

CONCLUSION

The conclusion summarizes the key findings from the comprehensive review and emphasizes the importance of continuous research and development efforts in performance enhancement and resource allocation in cloud computing. It highlights the need for interdisciplinary collaborations and further exploration of emerging technologies.

REFERENCES

1. Rosenthal A, Mork P, Li MH, Stanford J, Koester D, et al (2009). Cloud computing: a new business paradigm for biomedical information sharing. J Biomed Inform Apr. 43: 342–53.
2. Chenna R, Sugawara H, Koike T et al (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31: 3497–500.
3. Altschul SF, Lipman DJ (1990). Protein database searches for multiple alignments. Proc Natl Acad Sci USA. 87: 5509–5522.
4. Wall DP, Fraser HB, Hirsh AE (2011). Detecting putative orthologs. Bioinformatics. 19: 1710–1721.

5. Wall DP, Deluca T (2007). Ortholog detection using the reciprocal smallest distance algorithm. Methods Mol Biol.396: 95–110.

6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol. 215: 403–413.

7. Abdullah A, Ajay DK (2022). Cost-Effective Spot Instances Provisioning Using Features of Cloud Markets. Int J Cloud Comput. 12: 1-27.

8. Jouini M (2015). Mean Failure Cost Extension Model towards Security Threats Assessment: A Cloud Computing Case Study. J Comput. 10: 184-194.

9. Zhao S, Prenger K, Smith L, Messina T, Fan H (2013). Rainbow: a tool for large-scale whole-genome sequencing data analysis using cloud computing. BMC Genomics. 14: 425.

10. Banu SS, Balasundaram SR (2018).Cost Effective Approaches for Content Placement in Cloud CDN Using Dynamic Content Delivery Model. Int J Cloud Comput. 8: 78-117.