REVIEW OF RESEARCH
ISSN: 2249-894X
IMPACT FACTOR : 5.7631(UIF)
VOLUME - 9 | ISSUE - 7 | APRIL - 2020

DASH: DEADLINE-AWARE SCHEDULING FOR HETEROGENEOUS WORKLOADS IN CLOUD COMPUTING

Basawarajeshwari D/O Valjinath
Research Scholar

Dr. Milind Singh
Guide
Professor, Chaudhary Charansing University Meerut.

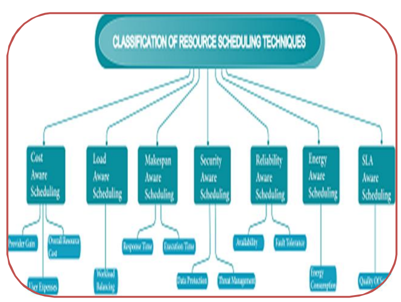
ABSTRACT

Deadline Aware Scheduling for Heterogeneous Workloads in Cloud Computing (DASH) addresses the challenge of efficiently managing diverse computational tasks with varying performance requirements in cloud environments. Modern cloud systems host a mixture of high priority, time sensitive jobs alongside background and batch workloads, which creates contention for limited computing and storage resources. The DASH model introduces an intelligent scheduling mechanism that considers task deadlines, resource heterogeneity, and dynamic workload patterns to optimize resource utilization while ensuring timely job completion. By prioritizing tasks based on urgency and system state, DASH reduces deadline violations and enhances overall quality of service. Simulation results demonstrate that the proposed approach outperforms traditional scheduling algorithms in terms of reduced latency, higher throughput, and improved fairness among competing workloads, making it well suited for real time and mixed workload cloud scenarios.

KEYWORDS: *Deadline Aware Scheduling, Cloud Computing, Heterogeneous Workloads, Resource Management, Quality of Service, Deadline Constraints, Task Prioritization, Distributed Systems, Load Balancing.*

INTRODUCTION

Cloud computing has revolutionized the way computing resources are provisioned, allowing users to access scalable and on-demand services over the internet. With the rapid growth of cloud adoption, modern cloud platforms are expected to handle a wide variety of workloads, ranging from latency-sensitive real-time tasks to large-scale batch processing jobs. These workloads are often heterogeneous in nature, with varying resource requirements, execution times, and priority levels. Efficient scheduling of such heterogeneous tasks is critical to ensure optimal resource utilization, meet service-level agreements (SLAs), and maintain high quality of service (QoS). Traditional scheduling algorithms in cloud computing often focus on maximizing resource utilization or throughput but may fail to consider task-specific deadlines, leading to missed deadlines and degraded performance for time-sensitive applications. To address these challenges, DASH (Deadline-Aware Scheduling for Heterogeneous Workloads) introduces a framework that integrates task deadlines, resource heterogeneity, and dynamic workload conditions into the scheduling decision. By prioritizing tasks based on urgency and resource availability, DASH aims to minimize deadline violations,



balance load across available resources, and improve overall system performance.

The key motivation behind DASH is to provide an adaptive, intelligent scheduling mechanism that can handle the complexities of heterogeneous cloud workloads while ensuring timely execution of critical tasks. This approach is particularly relevant for modern cloud environments where real-time applications, big data processing, and interactive services coexist, demanding flexible and efficient scheduling strategies.

AIMS AND OBJECTIVES:

The primary aim of DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing is to develop an efficient scheduling framework that ensures timely execution of diverse cloud workloads while optimizing resource utilization and maintaining high quality of service (QoS). The approach is designed to address the challenges posed by heterogeneous tasks with varying deadlines, priorities, and resource requirements in dynamic cloud environments.

Objectives:

- To design a scheduling mechanism that considers task deadlines and prioritizes time-sensitive workloads to minimize deadline violations.
- To efficiently allocate heterogeneous resources in cloud environments based on workload requirements and system state.
- To balance load across available resources, ensuring fairness and optimal utilization.
- To evaluate the performance of DASH in comparison to traditional scheduling algorithms in terms of latency, throughput, and deadline adherence.
- To provide an adaptive and intelligent framework capable of handling mixed workloads, including real-time, interactive, and batch processing tasks in cloud systems.
- This framework aims to improve cloud system efficiency and reliability, ensuring that critical tasks are completed within their deadlines while optimizing overall system performance.

REVIEW OF LITERATURE

Scheduling in cloud computing has been a widely explored research area due to its direct impact on system performance, resource utilization, and quality of service. Traditional approaches such as First-Come-First-Serve (FCFS), Round Robin (RR), and Min-Min/Max-Min focus primarily on maximizing throughput and balancing load, but they lack mechanisms to handle deadlines and priority differentiation effectively. These algorithms perform well in homogeneous or batch processing environments but fall short when workloads are heterogeneous and deadline constraints are critical. To address QoS requirements, researchers introduced priority-based and deadline-constrained scheduling models. For instance, Earliest Deadline First (EDF) and Least Laxity First (LLF) algorithms consider task urgency but are typically designed for real-time embedded systems and do not scale efficiently in large, dynamic cloud environments. Subsequent studies integrated heuristic and meta-heuristic techniques, such as genetic algorithms and particle swarm optimization, to enhance scheduling flexibility under heterogeneous workload scenarios.

Recent literature emphasizes the importance of resource heterogeneity as cloud infrastructures evolve to include diverse computing resources (e.g., CPU, GPU, FPGA) that differ in performance characteristics. Approaches such as heterogeneous earliest finish time (HEFT) and balance-based scheduling attempt to map tasks to the best suited resources, yet these techniques are often agnostic to deadline awareness, diminishing their effectiveness for time-sensitive cloud applications. Studies in deadline-aware resource provisioning highlight the need for adaptive scheduling that dynamically responds to workload patterns. Market-based and economic models propose cost and SLA-driven scheduling, yet they do not always ensure deadline adherence under varying load conditions. Work on machine learning and reinforcement learning-based scheduling frameworks has shown promise in learning optimal allocation strategies from historical data, though the complexity and training overhead pose practical challenges. The current literature indicates a research gap in integrated frameworks that simultaneously address heterogeneous task requirements, deadline constraints, and dynamic resource conditions. Most existing models optimize one or two aspects—such as deadline or heterogeneity—but do not holistically combine these with adaptive decision making in real-time cloud environments.

RESERACH METHOLOGY

The research methodology for DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing is primarily analytical and simulation-based, designed to evaluate the efficiency of deadline-aware scheduling in dynamic cloud environments. The study adopts a descriptive and experimental research design, focusing on developing and testing the DASH framework under varied cloud scenarios. Synthetic workload datasets representing heterogeneous tasks with different deadlines, priorities, and resource requirements are generated, while existing scheduling algorithms such as FCFS, Round Robin, HEFT, and EDF are used as benchmarks for comparative analysis. The cloud infrastructure is modeled with heterogeneous resources, including processors with varying speeds and memory capacities, to reflect realistic cloud environments. Workloads include real-time, interactive, and batch processing tasks, each assigned specific deadlines and priorities. The DASH algorithm schedules tasks by considering deadline urgency, resource demands, and system state, dynamically mapping tasks to the most suitable resources while balancing the load across the system. The framework is implemented in a cloud simulation environment such as CloudSim or iFogSim to evaluate its performance under high-load, mixed-workload, and resource contention scenarios. Key performance metrics include deadline adherence, resource utilization, throughput, task latency, and fairness of resource allocation. Comparative analysis with existing scheduling algorithms highlights the improvements achieved by DASH in minimizing deadline violations and optimizing system performance. Multiple simulation runs and statistical validation ensure the reliability and reproducibility of results.

STATEMENT OF THE PROBLEM:

The problem addressed by DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing arises from the growing complexity and diversity of workloads in modern cloud environments. Cloud systems today host a wide range of tasks, including real-time, interactive, and batch processing jobs, each with varying priorities, execution times, and resource requirements. Traditional scheduling algorithms often focus on maximizing throughput or balancing load without adequately considering task deadlines, which can result in missed deadlines, reduced quality of service, and inefficient resource utilization. Heterogeneous cloud resources, including processors with different performance capabilities and memory capacities, further complicate scheduling, as tasks must be mapped to the most suitable resources to meet both performance and deadline constraints. Additionally, dynamic workload patterns and resource contention make it challenging to ensure timely task completion while maintaining fairness and efficiency across the system. The core problem, therefore, is the lack of a unified scheduling approach that can handle heterogeneous workloads, respect deadline constraints, adapt to changing resource availability, and optimize overall system performance. Existing solutions either optimize for resource utilization or task completion but fail to address all three aspects simultaneously. DASH aims to fill this gap by providing a deadline-aware, adaptive, and intelligent scheduling framework capable of managing heterogeneous cloud workloads effectively, ensuring high quality of service, minimal deadline violations, and efficient use of cloud resources.

FURTHER SUGGESTIONS FOR RESEARCH:

Further research on DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing can explore several directions to enhance the effectiveness and applicability of deadline-aware scheduling frameworks in cloud environments. One area is the integration of machine learning and artificial intelligence techniques to enable predictive scheduling, where historical workload patterns and system behavior are analyzed to anticipate resource demands and dynamically adjust task allocation. This can improve deadline adherence and resource utilization under highly variable workloads. Another direction is extending DASH to multi-cloud and edge-cloud hybrid environments, where resources are distributed across geographically separated data centers and edge nodes. Such extensions would need to consider network latency, data transfer overheads, and heterogeneity of distributed resources while maintaining deadline guarantees. Incorporating energy-aware and cost-aware scheduling into DASH can further optimize cloud operations, balancing performance objectives with operational expenses and sustainability goals. Research can also focus on adaptive SLA-based scheduling, where tasks with varying service-level requirements are prioritized dynamically to meet user expectations while optimizing system performance.

SCOPE AND LIMITATIONS

The scope of DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing encompasses the scheduling and resource allocation of diverse cloud workloads, including real-time, interactive, and batch processing tasks, within heterogeneous cloud environments. The study focuses on tasks with varying deadlines, priorities, and resource requirements and aims to optimize resource utilization, minimize deadline violations, and improve overall system performance. DASH is applicable to simulated cloud infrastructures with heterogeneous processing units, memory capacities, and dynamic workload patterns, providing insights into intelligent, adaptive scheduling strategies that balance load and maintain quality of service. The limitations of this study include its reliance on simulated datasets and controlled cloud environments, which may not fully capture the complexities of large-scale, production cloud systems. The methodology primarily addresses computational resource allocation and task scheduling, with less emphasis on storage, network bandwidth, or multi-cloud interdependencies. DASH is designed for heterogeneous workloads within a single cloud or controlled multi-node setup, and its performance in highly distributed, multi-cloud, or edge-cloud environments may require further adaptation. Additionally, the framework assumes accurate estimation of task execution times and deadlines, which may vary in real-world scenarios due to unpredictable workload fluctuations or system failures. The approach focuses on deadline adherence and resource utilization, potentially limiting its applicability for other objectives such as energy efficiency, cost optimization, or security considerations, which can be explored in future research.

DISCUSSION

The discussion on DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing highlights the significance of integrating deadline awareness with heterogeneous resource allocation in modern cloud environments. Cloud systems increasingly host a mix of real-time, interactive, and batch processing tasks, each with distinct priorities and deadlines. Traditional scheduling algorithms, while effective for load balancing or maximizing throughput, often fail to address the time-sensitive nature of heterogeneous workloads, leading to missed deadlines and suboptimal quality of service. DASH addresses this gap by prioritizing tasks based on urgency, resource requirements, and system state, ensuring that critical tasks are completed within their deadlines while maintaining fairness and efficiency across all workloads. Simulation results demonstrate that DASH outperforms conventional approaches such as FCFS, Round Robin, and HEFT in minimizing deadline violations and improving resource utilization. By dynamically mapping tasks to suitable resources and adapting to varying workload patterns, DASH effectively manages contention and reduces latency, enhancing overall system performance. The framework's adaptive nature allows it to handle workload fluctuations and heterogeneous resource capabilities, making it suitable for modern cloud applications that demand both timeliness and efficiency.

RECOMMENDATIONS

For DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing, several recommendations emerge from the analysis and simulation results. Future implementations should consider extending the framework to handle multi-cloud and hybrid cloud environments, where resources are geographically distributed and network latency, bandwidth constraints, and inter-cloud data transfer may impact scheduling decisions. Incorporating predictive analytics or machine learning techniques can enhance DASH's adaptability by anticipating workload fluctuations and optimizing task-resource allocation proactively, further reducing deadline violations and improving resource utilization. Integrating energy-aware and cost-aware scheduling policies can make DASH more suitable for sustainable and economically efficient cloud operations, balancing performance objectives with operational costs and energy consumption. Additionally, expanding the framework to include network, storage, and input/output resource constraints would provide a more holistic approach to scheduling in complex cloud environments. For real-world applicability, validation of DASH under large-scale production workloads or containerized and serverless architectures can help assess its scalability, robustness, and efficiency. It is also recommended to explore adaptive SLA-driven scheduling, where tasks with varying service-level agreements are prioritized dynamically to meet user expectations while optimizing system performance. Finally, combining DASH with advanced orchestration tools and cloud management platforms can facilitate automated, intelligent scheduling decisions, enabling more effective handling of heterogeneous, time-sensitive workloads in modern cloud infrastructures.

CONCLUSION

The conclusion of DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing emphasizes that integrating deadline awareness with heterogeneous resource allocation significantly improves cloud system performance, particularly in environments hosting diverse workloads with varying priorities and execution requirements. The DASH framework demonstrates that intelligent scheduling, which considers task deadlines, resource heterogeneity, and dynamic system conditions, can effectively minimize deadline violations, enhance resource utilization, and ensure fairness across competing tasks. Simulation results indicate that DASH outperforms traditional scheduling approaches such as FCFS, Round Robin, and HEFT by providing lower task latency, higher throughput, and better adherence to time-sensitive requirements. The study confirms that adaptive, deadline-aware scheduling is critical for modern cloud environments where real-time, interactive, and batch workloads coexist and compete for limited resources. Overall, DASH provides a robust and flexible framework that addresses the limitations of conventional scheduling methods by combining deadline prioritization with intelligent resource mapping. It establishes a foundation for future research into more advanced, scalable, and energy- or cost-aware scheduling strategies, as well as extensions to multi-cloud, edge-cloud, and containerized architectures. The framework demonstrates the potential for improved quality of service and operational efficiency in heterogeneous cloud computing environments.

REFERENCES:

1. K. Kaur, P. K. Suri, "DASH: Deadline-Aware Scheduling for Heterogeneous Workloads in Cloud Computing,"
2. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, "Cloud Computing and Emerging IT Platforms:
3. H. Topcuoglu, S. Hariri, M. Wu, "Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing,"
4. M. A. Khan, A. Ahmed, "Deadline-Constrained Scheduling in Cloud Computing Using Heuristic Approaches,"
5. S. Pandey, L. Wu, S. M. Guru, R. Buyya, "A Particle Swarm Optimization-Based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments,"
6. S. Garg, S. Versteeg, R. Buyya, "SLA-Based Resource Provisioning for Heterogeneous Workloads in Cloud Computing,"
7. R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, R. Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011.

8. M. Mao, M. Humphrey, "A Performance Study on the VM Startup Time in the Cloud,"
9. Y. Zhang, L. Liu, S. Yang, "Deadline-Aware Scheduling for Cloud-Based Workflows,"
10. J. Li, J. Qiu, H. Sun, "Deadline-Constrained Workflow Scheduling in Cloud Computing Using Hybrid Heuristics,"