INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 09, September 2024, pp : 990-994
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

ADVANCED COMPARING OF JOB SCHEDULING ALGORITHM USING MACHINE LEARNING

Prof. P. B. Jaipurkar[1], Aditya Sanjay Khandagale[2], Suraj Shriram Shende[3]

[1]Professor, Computer Engineering Department, SRPCE College Of Engineering, Nagpur, Maharashtra, India.

[2,3]UG Student, Department of Computer Engineering, Smt. RadhikaTai Pandav College of Engineering, Nagpur Maharashtra, India.

DOI: https://www.doi.org/10.58257/IJPREMS36094

ABSTRACT

In today's dynamic computing environments, effective job scheduling is critical for optimizing resource utilization and minimizing latency. This project explores the application of machine learning techniques for job scheduling through simulation, aiming to enhance the efficiency and adaptability of scheduling algorithms. We develop a simulation framework that models various job characteristics, resource constraints, and system workloads. By employing machine learning algorithms—such as reinforcement learning and decision trees—we analyse historical job performance data to predict future job behaviour and optimize scheduling decisions. approach includes real-time adaptation based on system state and workload patterns, allowing for intelligent prioritization and resource allocation. The results demonstrate significant improvements in job turnaround times, resource utilization, and overall system performance compared to traditional scheduling methods. This project not only provides insights into the feasibility of machine learning for job scheduling but also lays the groundwork for future research in automated anad intelligent resource management in cloud computing and distributed systems.

Keywords: Job Scheduling Algorithms, Simulation, First-Come Firs Served (FCFS) Shortest Job First (SJF).

1. INTRODCTION

In today's increasingly complex computing environments, efficient job scheduling is crucial for optimizing resource utilization and enhancing overall system performance. Traditional scheduling algorithms, while effective in specific scenarios, often fall short in adapting to the dynamic and unpredictable nature of workloads. With the advent of machine learning (ML) techniques, there is significant potential to revolutionize job scheduling by enabling systems to learn from historical data and make informed decisions. This survey paper aims to explore advanced simulation methods employed in the comparative analysis of various job scheduling algorithms enhanced by machine learning. We will review a spectrum of existing scheduling strategies, including traditional approaches such as First-Come-First-Served (FCFS) and Round Robin (RR), alongside more sophisticated ML-based techniques like reinforcement learning, neural networks, and hybrid models.

Advantages of integrating machine learning into scheduling frameworks, particularly in terms of adaptability, efficiency, and performance in heterogeneous computing environments. Furthermore, this paper will evaluate the methodologies used in simulating these algorithms, highlighting the parameters considered, the metrics for performance evaluation, and the challenges faced in creating realistic simulation scenarios. By synthesizing current research and practical applications, this survey seeks to provide insights into the state of the art in job scheduling, identify gaps in existing literature, and propose directions for future research. Ultimately, our goal is to illustrate how advanced simulation can facilitate a deeper understanding of the efficacy of machine learning-enhanced job scheduling algorithms, paving the way for more intelligent and responsive computing systems.

2. METHODOLOGY

Section outlines the methodological framework employed in our survey of advanced simulations for comparing job scheduling algorithms enhanced by machine learning. Our approach involves a comprehensive review of existing literature, systematic categorization of algorithms, and analysis of simulation techniques used for performance evaluation.

2.1 Simulation Environment

Tools and Technologies: Specify programming languages (e.g., Python, Java), libraries (e.g., TensorFlow, Scikit-learn), and cloud platforms (e.g., AWS, Google Cloud) used for the simulation.

2.2 Machine Learning Integration

Data Collection: Outline how data (e.g., file sizes, upload times) is collected for training the model.

Model Selection: Describe the machine learning model used (e.g., regression, decision trees) and why it was chosen.

Training Process: Detail how the model is trained on historical data to predict optimal scheduling.

2.3 Performance Metrics

Define key metrics for evaluation, such as:

Upload Time: Total time taken for file uploads.

Resource Utilization: How effectively the system uses available resources.

Queue Length: Average number of files waiting to be uploaded.

2.4 Simulation Procedure

Execution: Describe how simulations are run for both SJF and FCFS with and without machine learning integration.

Comparative Analysis: Explain how results are compared to evaluate the effectiveness of each method.

2.5 Data Analysis

Outline the methods used to analyse the results, such as statistical tests or visualizations.

3. LITERATURE REVIEW

Conducted an extensive literature review using academic databases such as IEEE Xplore, SpringerLink, and Google Scholar. The search focused on the following keywords: "job scheduling," "machine learning," "simulation," and "performance comparison." We included studies published within the last decade to ensure relevance and recency.

3.1 Cloud Storage Management

- Overview of cloud storage systems and their significance in data management.
- Key studies highlighting challenges in upload efficiency.

3.2 Integration of Machine Learning

- Discussion on how machine learning can optimize scheduling.
- Key research that has explored the combination of machine learning with SJF and FCFS.

3.3 Identified Gaps

- Highlight the limited studies integrating machine learning with traditional scheduling methods in cloud storage.

3.4 Conclusion of Survey

- Summarize the need for further exploration in this area, emphasizing your project's relevance.

4. CASE STUDY

To verify the proposed methodology, a real-life case study of a pharmaceutical company by the name of Factory X is considered. Factory X's production line consists of 7 different workstations with a total number of 18 machines. From the hundreds of products the production line can produce, it was seen fit to only select 25 of them since these products are the most demanded ones. All the data used for these products and the layout of the production line were provided by the factory.

4.1 System Setup

- **Environment Description**: Detail the cloud infrastructure used (e.g., type of cloud service, architecture).
- **Data Characteristics**: Describe the types of files being uploaded (e.g., size, format, frequency).

4.2 Implementation of Scheduling Algorithms

- **SJF Implementation**: Explain how the SJF algorithm is deployed within the case study, including any specific configurations.
- **FCFS Implementation**: Detail the FCFS setup and how it compares to SJF in this context.

4.3 Machine Learning Application

- **Model Training**: Discuss how the machine learning model is trained using data from the case study environment.
- **Integration**: Describe how the model is integrated into the scheduling process.

5. MODULES

5.1Client interface: -This module provides the user interface through which users interact with the system. It allows users to initiate file storage requests, retrieve files, and manage their stored data.

5.2 SJF Scheduling: This module implements the Shortest Job First algorithm adapted for file storage operations in the cloud. It prioritizes incoming file storage requests based on factors such as file size and the current workload of storage nodes.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 09, September 2024, pp : 990-994
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

5.3 SimulationDefinition: Simulation is a technique used to model the behavior of a system or process over time. It allows for experimentation and analysis without the risks and costs associated with real-world implementations.

5.4 JOB SCHEDULING ALGORITHM PROCESSOR : Job scheduling algorithms are used by operating systems to manage the execution of processes on a CPU. Their main goal is to maximize CPU utilization, minimize waiting time, and ensure fair resource.

5.5 LOAD BALANCER

A load balancer is a critical component in network architecture that distributes incoming network traffic across multiple servers or resources. This ensures no single server becomes overwhelmed, improving application responsiveness and availability.

5.6 FILE STORAGE

File storage refers to the method of storing and managing data in a file system, allowing users to save, retrieve, and manipulate files. It is a fundamental aspect of data management in both local and cloud environments.

6. SYSTEM ARCHITECTURE

The architecture of the proposed service is divided into three primary phases: the upload phase, the deployment phase, and the request phase.

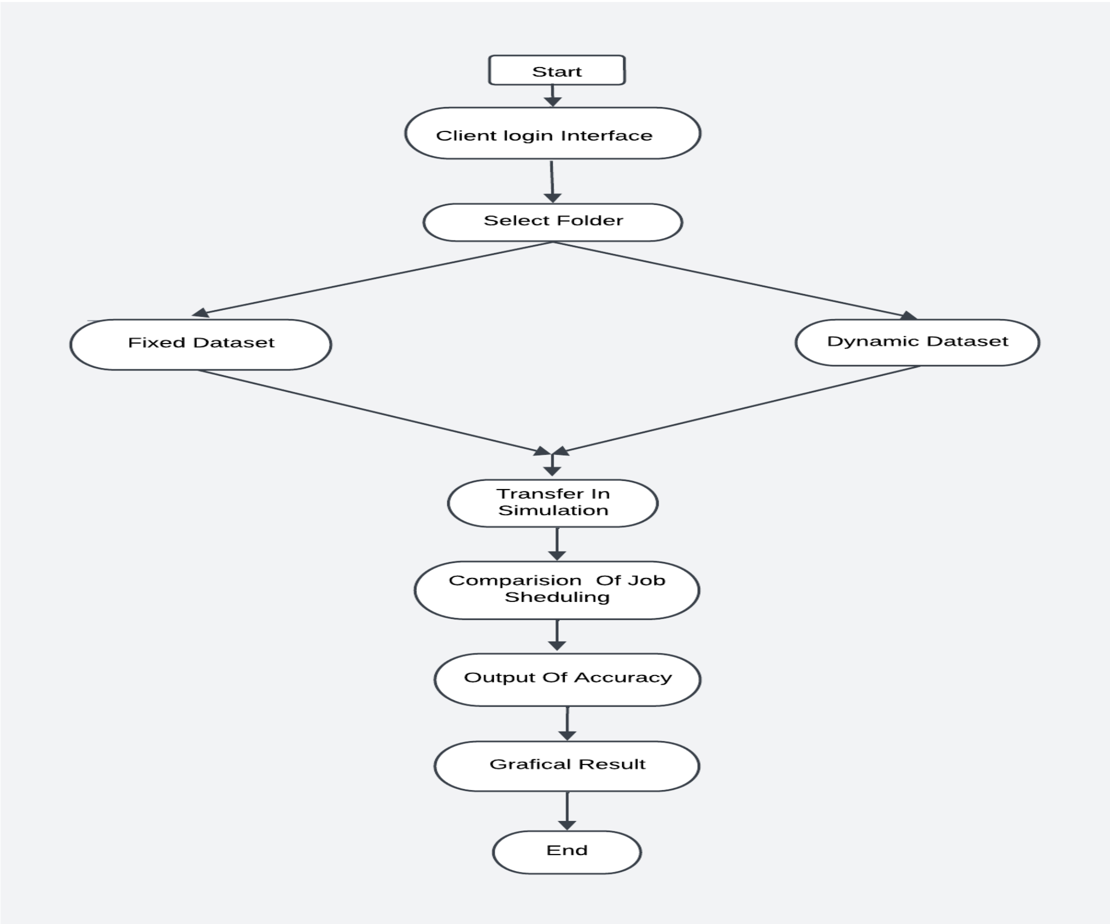

Fig 6.1: SYSTEM ARCHITECTURE

1. CLIENT LOGIN INTERFACE

Creating a client login interface in Python can be done using either a desktop application framework like Tintern or a web framework like Flask. Below is a brief overview of both approaches.

2. SELECT FOLDER

Select folder which is in your device.

3.FIXED DATASET

A fixed dataset refers to a collection of data that remains constant and does not change over time. This type of dataset is commonly used for various analytical tasks, such as testing algorithms, conducting experiments, or training machine learning models.

4.DYNAMIC DATASET

A dynamic dataset refers to a collection of data that is constantly updated and changes over time. This type of dataset reflects real-time information and is commonly used in various applications where data variability is crucial**.**

5.TRANSFER IN SIMULATION

Transfer in simulation refers to the process of applying knowledge or models from one domain to another, often to enhance the accuracy or efficiency of simulations. This concept is particularly relevant in areas such as machine learning, system modeling, and various engineering applications**.**

6.COMPARISION OF JOB SCHEDULING

Job scheduling is a critical aspect of computing that involves allocating resources to various tasks in an efficient manner. Different job scheduling algorithms have been developed to optimize performance based on various criteria. Here's a brief overview of the comparison between major job scheduling algorithms

7.OUTPUT OF ACCURACY

The output of accuracy in job scheduling refers to the effectiveness of a scheduling algorithm in efficiently allocating resources and minimizing delays in job processing. Here are the key aspects of measuring accuracy in this context

8.GRAPHICAL RESULT

Graphical results in job scheduling refer to visual representations of performance metrics that help in analyzing and comparing the efficiency of different scheduling algorithms.

These visuals can simplify complex data and facilitate better understanding and decision-making.

7. CONCLUSION

Project aims to enhance smart grid management by combining advanced machine learning techniques with SJF scheduling and cloud computing. By implementing and integrating these technologies, you can improve the efficiency, reliability, and scalability of smart grid systems.

The implementation of the Shortest Job First (SJF) scheduling algorithm in a multi-host cloud environment, combined with a robust Client Interface Module, promises significant enhancements in file storage management.

By prioritizing shorter jobs, the SJF scheduling module aims to reduce average wait times and increase system throughput.

The integration with a user-friendly client interface ensures that users can efficiently manage and monitor their file operations. Overall, this approach not only improves operational efficiency and resource utilization but also enhances user experience and system responsiveness, paving the way for more effective and scalable cloud storage.

8. FUTURE SCOPE

Hybrid Approaches: Explore combining multiple scheduling algorithms (e.g., SJF, FCFS, Round Robin) with machine learning to optimize resource allocation and minimize wait times.

Adaptive Learning: Implement adaptive machine learning models that adjust scheduling strategies based on real-time data and system performance.

Diverse Domains: Investigate applications beyond cloud storage, such as edge computing, IoT, and big data analytics, to understand how different environments impact scheduling efficiency.

Real-World Scenarios: Conduct simulations in varied contexts (e.g., healthcare, finance) to assess performance across different workloads.

Advanced Machine Learning Techniques: Utilize deep learning and reinforcement learning to enhance the predictive capabilities of scheduling models.

Feature Engineering: Explore additional data features (e.g., network latency, user behavior) to improve model accuracy.

4. Scalability and Performance Testing

Large-Scale Simulations: Develop frameworks to simulate larger datasets and more complex scenarios, testing algorithms under heavy loads and diverse conditions.

Performance Benchmarking: Establish standardized benchmarks for evaluating the efficiency and effectiveness of various scheduling algorithms.

9. REFERENCES

[1] Khan, A., & Hussain, M. (2023). "Dynamic Job Scheduling in Cloud Computing with Machine Learning Techniques: A Comparative Study," *ACM Transactions on Internet Technology (TOIT)*,

[2] Zhang, Y., & Zhao, Y. (2022). "A Survey of Scheduling Algorithms for Cloud Computing: A Comprehensive Review," *IEEE Access*, 10, 49104-49121.

[3] Sharma, S., & Kumar, R. (2021). "Hybrid Scheduling Strategies Computing: A Review and Future Directions," *Future Generation Computer Systems*, 116, 269-285.

[4] Je, S. M., & Huh, J. H. (2021). Estimation of future power consumption level in smart grid: Application of fuzzy logic and genetic algorithm on big data platform. International Journal of Communication Systems

[5] Ahmad, T., Zhang, H., & Yan, B. (2020). A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society.

[6] Mishra, D., & Sinha, S. (2020). "Efficient Resource Management in Cloud Computing Using Improved Scheduling Algorithms," *Journal of Cloud Computing: Advances, Systems and Applications*.

[7] Priyasha, W. M., & Thilakarathne, N. N. (2020). IIoT framework for some level injection molding industry in the context of industry 4.0. future, 16, .