ABSTRACT

This thesis delves into various performance improvement techniques for multi-core clusters, which are critical for advancing High-Performance Computing (HPC). Multi-core clusters are fundamental in supporting complex computational tasks, and optimizing their performance can significantly enhance computational efficiency and system reliability.

The research begins by identifying key challenges in the HPC environment, particularly focusing on memory congestion, resource prediction, and load balancing. It introduces an innovative scheme designed to mitigate memory congestion, thereby improving data throughput in multi-core systems. This scheme employs advanced parallel data mining techniques on public datasets, allowing for the development of predictive models that can forecast CPU performance. These models are instrumental in facilitating more efficient scheduling and resource allocation.

A significant portion of the thesis is dedicated to exploring load balancing mechanisms. It evaluates various algorithms and their effectiveness in heterogeneous cloud environments, offering a comprehensive survey of existing techniques and proposing new methodologies. The proposed solutions are validated through extensive simulations and real-world applications, demonstrating substantial improvements in computational efficiency.

Additionally, the research addresses dependency prediction of long-term resource usage in HPC environments. By employing machine learning techniques, the study develops models that can predict resource usage patterns, thus enabling better resource management and allocation strategies.

The thesis also examines the impact of different scheduling algorithms on system performance. It introduces an effective scheme for memory congestion reduction and

evaluates its performance through detailed experimental analysis. The results indicate a significant reduction in memory congestion, leading to enhanced system performance and reliability.

Moreover, the research explores the integration of advanced load balancing techniques with existing HPC infrastructures. It proposes a hybrid approach that combines multiple algorithms to optimize load distribution across multi-core clusters. This approach is shown to improve overall system performance, reduce processing time, and enhance system scalability.

In conclusion, this thesis makes significant contributions to the field of HPC by providing scalable and adaptive methods for improving the performance of multi-core clusters. The findings pave the way for the development of more robust and efficient computing infrastructures, which are essential for meeting the growing demands of complex computational tasks. These advancements are expected to have a profound impact on the future of HPC, enabling more efficient and reliable computational environments.