World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:17, No:03, 2023

Machine Learning Assisted Performance Optimization in Memory Tiering

Authors : Derssie Mebratu

Abstract : As a large variety of micro services, web services, social graphic applications, and media applications are continuously developed, it is substantially vital to design and build a reliable, efficient, and faster memory tiering system. Despite limited design, implementation, and deployment in the last few years, several techniques are currently developed to improve a memory tiering system in a cloud. Some of these techniques are to develop an optimal scanning frequency; improve and track pages movement; identify pages that recently accessed; store pages across each tiering, and then identify pages as a hot, warm, and cold so that hot pages can store in the first tiering Dynamic Random Access Memory (DRAM) and warm pages store in the second tiering Compute Express Link(CXL) and cold pages store in the third tiering Non-Volatile Memory (NVM). Apart from the current proposal and implementation, we also develop a new technique based on a machine learning algorithm in that the throughput produced 25% improved performance compared to the performance produced by the baseline as well as the latency produced 95% improved performance compared to the performance produced by the baseline.

Keywords : machine learning, bayesian optimization, memory tiering, CXL, DRAM
Conference Title : ICCAD 2023 : International Conference on Computer Architecture and Design
Conference Location : Dubai, United Arab Emirates
Conference Dates : March 16-17, 2023