CS446 Project Proposal: HPMMAP Implementation for Databases
Author: Ishaan Narain
Date: 17/04/2023

Project Introduction

        HPMMAP (High Performance Memory Mapping and Allocation Platform) is a lightweight memory subsystem for High Performance Computing (HPC) applications in a commodity OS/R environment, based on the paper written by Kocoloski et al, 2014. The goal of the subsystem is to optimize application performance and decrease variance by introducing a transparent and configurable layer without any overhead, achieving a 50% decrease in runtime.

        The overarching goal of this project is to integrate the HPMMAP code (from 2015) into a modern linux system and test the performance of the subsystem on a database system that applies the MMAP() system call, such as LMDB, SQLite, or WiredTiger. The project will be carried out in 3 steps. Step 1 includes porting and running the HPMMAP code into a modern Linux environment, step 2 involves integrating HPMMAP with a database engine, such as the ones listed above, and step 3 involves testing the performance of an HPMMAP system and comparing it to the regular MMAP system call for a database application.

Project Motivations

        The reason for choosing this project is due to my interest in low level software development for finance / AI systems which rely on high performance computing systems. The project will involve a deep dive into porting kernel code bases for modern linux, implementing source code with various applications like databases, and setting up tools to evaluate and measure the performance.

Objectives and Deliverables

        The primary objective of this project is to provide a codebase with a functional HPMMAP in a modern linux environment, as well as the source code for the implementation of HPMMAP in a database environment. The secondary objective is to test the implementation of HPMMAP in a database environment and collect the performance results to compare against a vanilla MMAP implementation in a Linux environment. The deliverables of this project will contribute to the field of High Performance Computing and Database applications by providing a modern custom subsystem to the memory management system on Linux Operating Systems and providing practical examples of use cases with performance data.

The results of this project will be summarized in a research paper and presentation, encapsulating the code base and results of the performance experiments.

Timelines (Probably aggressive)

Week 4 (17/04/2023 - 23/04/2023)
- Setup functional codebase for development and testing with QEMU and HPMMAP
- Get familiar with the HPMMAP codebase
- Begin porting over the HPMMAP database for current Linux

Week 5 (24/04/2023 - 30/04/2023)
- Continue to work on porting the HPMMAP codebase

Week 6 (01/05/2023 - 07/05/2023)
- Begin implementing the HPMMAP codebase in a database environment
- Begin working on a paper process summarizing the port of HPMMAP

Week 7 (08/05/2023 - 14/05/2023)
- Complete the implementation of the HPMMAP codebase and start the paper process

Week 8 (15/05/2023 - 21/05/2023)
- Complete testing and collection of performance results in the database environment
- Begin working on a paper for the performance results

Week 9 (22/05/2023 - 28/05/2023)
- Wrap up any other implementation issues
- Complete the paper and a presentation summary of the results

Week 10 (29/05/2023 - 04/06/2023)
- Wrap up any other implementation issues
- Final draft of the paper and presentation

References

[1] B. Kocoloski and J. Lange, "HPMMAP: Lightweight Memory Management for Commodity Operating Systems," *2014 IEEE 28th International Parallel and Distributed Processing Symposium*, Phoenix, AZ, USA, 2014, pp. 649-658, doi: 10.1109/IPDPS.2014.73.

[2] Kocoloski Brian, 2015, HPMMAP, https://gitlab.prognosticlab.org/prognosticlab/hpmmap