

IEEE International Symposium on Workload Characterization – 2022

Title: Methods for characterizing workloads with Hardware accelerated memory page compression

Authors: Ravindran, Binuraj; Al-Fahad, Rakib; Liang, Dan; Chowdhury, Muktadir; Luo, Zhenlin, Intel Corporation

Abstract:

With DRAM demand ever increasing with emerging data center workloads, DRAM cost significantly contributes to the data center infrastructure cost. Cloud service providers have already deployed memory-page-compression techniques to reduce the effective memory footprint of workloads without compromising too much on the workload performance. Besides the state-of-the-art software compression, hardware accelerated memory-page compression can lower the overhead of doing compression/decompression and allow parallel processing. Thus, this method can further reduce the memory footprint and DRAM costs. An effective methodology to characterize the workloads is necessary to understand the effectiveness of memory page compression. This paper covers a few techniques to characterize workloads in terms of compressibility, page fault latency, and memory savings when deploying hardware accelerated memory page compression. It describes how the methodology can be extended using SPECPU as a proxy workload,

The main topics that will be covered are:

- Challenges in characterizing workloads and accurately reporting memory savings, page-fault latency, and compression ratio
- Dynamic Squeezer – A tool for effectively reducing memory usage and monitoring memory-savings
- Examples with SpecCPU as a proxy-workload
- How a hardware accelerator for compression/decompression improves memory savings?