Data Management in Heterogeneous Memory Systems

Thaleia Dimitra Doudali, Ada Gavrilovska
thdoudali@gatech.edu, ada@cc.gatech.edu

1. Motivation

- High Performance Computing and Big Data applications have dataset sizes that often exceed the available DRAM capacities.
- Emerging memory technologies that are much cheaper, such as Non Volatile Memory, are used to extend the memory space creating a heterogeneous memory subsystem.
- Data in Non Volatile Memory will incur higher access latencies, affecting the application performance, slowing it down compared to an ideal case when all data could fit in DRAM.
- Existing solutions reduce the performance slowdown by prioritizing allocations of the most frequently accessed objects in DRAM. However, they assume fixed hardware capacities.

2. Problem Statement

![Big Data Applications](facebook) ![HPC Applications](twitter) DRAM NVM SLOW

Problem: How do we size the memory components and manage data over the heterogeneous memory system, so as to minimize the application performance slowdown?

3. Observations

- Not all applications are slowed down in the same degree when accessing Non Volatile Memory.
- Not all data objects of an application help reduce the performance slowdown when allocated in DRAM.

4. CoMerge Solution

**CoMerge:** Memory sharing policy that prioritizes DRAM allocations for critical data objects. Achieves:
- Lower runtime across all collocated applications.
- Higher DRAM utilization.

![Equal Split](xsbench) ![unused](clomp) ![stream] CoMerge 7x slowdown 2.6x

5. Mnemo Solution

**Mnemo:** Profiling tool that estimates the application performance slowdown for incremental DRAM capacity on a heterogeneous memory system.

![Key COO](Key KO COO) ![< 1% accuracy error](watermark) Runtime slowdown (%) vs. Keys in DRAM 0 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 Keys in DRAM

6. Future Directions

- **Mnemo**
- **Use Cases**
- **Applications**

Goal: connect the rest of the dots.

All work is part of the projects:
- ECP SICM (Software Interface to Complex Memories)
- SSIO Unity (Unified Memory and Storage Space)