Aiman Priester

P10: Reducing DRAM Footprint with NVM in Facebook

Summary

The paper studies a system that incorporates properties of DRAM and NVM to create a system called MyNVM. MyNVM, built on top of MyRocks, aims to reduce the total cost of ownership (TCO) for data centers. The paper cites that due to scaling problems of DRAM (also talked about in P09), TCO has increased substantially as data centers store more data. Data must still be accessed in a short amount of time. While NVM by itself is not a prime solution due to its low read bandwidth, using the technology as an L2 Cache will reduce the read times as NVM is still substantially faster than flash storage. By partitioning database indices, DRAM space can be freed up for "hotter" data blocks and NVM can be used to store the cooler blocks. Of course, flash storage will still be used as main memory if L1 and L2 caches miss. This results in the reduced dependence of DRAM, enabling data centers to lower their TCO.

Strengths

Essentially, this paper dives into solving NVM limitations. While NVM has a higher latency than DRAM, it compensates by being much more cost effective on a per-byte basis. The paper cites replacing 80GB of DRAM with 140GB of NVM, which increased the hit rate for the L2 cache. The authors have also provided a solution to NVMs inherent low durability by maintaining an LRU within DRAM as part of an High Hit Rate Admission Policy. This policy enables the NVM a reduced chance of hitting its Daily Write Per Day (DWPP) limit, furthering its lifetime.
Another provision that the authors came up with is that a miss in L2 (NVM) does not necessarily cause a load from flash. An NVM Lookup Table is stored in DRAM and is used to store a hashed signature, to the tune of 4 per block. If the data is deemed "hot", then said data will be written to NVM in a tune of at least 128KB. The incorporation of NVM and DRAM, with limitations of NVM padded, results in performance comparable to MyRocks with high DRAM.

It should also be noted that the authors also opted to make a switch from Interrupt handling to polling. Replacing interrupts allowed reduced CPU consumption and increased NVM bandwidth for transferring data blocks from flash.

Weaknesses

For the most part, the author's implementation will cause a speed up on the average case. However, anytime an additional layer is added to the memory hierarchy, the worst case scenario will be in the event of consecutive misses in both L1 and L2, the total time will increase as it has to traverse one additional level. While this is unlikely to happen as time increases, it should still be considered in the papers findings / conclusions.

Unresolved Issues

While it is true that on paper it will be much cheaper to implement MyNVM, the paper does not consider the time and effort required to pair and additional level of cache. One could argue that it would make sense if the physical storage device is merged on a hardware level. However, it is unclear weather on-paper cost benefits will outweigh total implementation cost when porting to the real world.

Discussion

The costs cited in the paper are consumer prices. It is more than likely that large companies such as Facebook will get a significantly reduced price on hardware used. With that in mind, wouldn't the final cost difference of this implementation be negligible, especially for a company with a substantial market cap?