

Data Migration Policies in a Copy-on-Write Tiered Storage Stack - Conception and Implementation

- ▶ Author: Johannes Wünsche
- ▶ Type: Master's Thesis
- ▶ Date: 2022-10-17
- ▶ Reviewers: Jun.-Prof. Dr. Michael Kuhn, Dr. David Broneske
- ▶ Supervisors: Jun.-Prof. Dr. Michael Kuhn
- ▶ Download: PDF

We currently observe a trend of ever-increasing data demand in a range of computation domains, these include high-performance computing, databases, and many more. Satisfying these growing storage requirements, the continuous improvement and scaling out of existing technology have been furthered immensely in the last decade, as well as the introduction of new mediums and interfaces. Varying advantages have emerged for these mediums. To utilize them, different storage mediums are combined in heterogeneous storage systems, which aim to optimize the usage of storage mediums to prevalent data flows, such as burst-write patterns. However, the organization of these systems is an NP-hard problem, which is, furthermore, extraordinarily hard to approximate. Hence, so-called "migration" policies, act on proxy values such as data temperature, access frequency, or access recency to gain improvements over the status quo. Additionally, with concepts like Copy on Write and write-optimized data structures becoming more common, held assumptions about data distribution and latency have changed. We explore policies and issues of data migration in a hierarchical heterogeneous storage stack based on write-optimized B-epsilon-trees by implementing migration functionality and a dynamic policy interface with two proven policies from other papers as example implementations. We find that strong differences in placement exist, based on data size and access type, contradicting usual assumptions made on non-write-optimized storage. Furthermore, when observing I/O time, we see a 30% speedup in common write patterns, with remaining potential.