

## **Read-Aware LSM for Disaggregated Heterogeneous Storage**

In recent years, emerging hardware storage technologies have focused on divergent goals: better performance or lower cost- per-bit of storage. Correspondingly, data systems that employ these new technologies are optimized either to be fast (but expensive) or cheap (but slow). We take a different approach: by combining multiple tiers of fast and low-cost storage technologies within the same system, we can achieve a Pareto-efficient balance between performance and cost-per-bit.

This paper presents the design and implementation of PrismDB, a novel log-structured merge tree based key-value store that exploits a full spectrum of heterogeneous storage technologies (from 3D XPoint to QLC NAND). We introduce the notion of “read-awareness” to log-structured merge trees, which allows hot objects to be pinned to faster storage, achieving better tiering and hot-cold separation of objects. Compared to the standard use of RocksDB on flash in datacenters today, PrismDB’s average throughput on heterogeneous storage is 2.3x faster and its tail latency is more than an order of magnitude better, using hardware that is half the cost.