

## **Data Similarity-Aware Computation Infrastructure for the Cloud**

### **Abstract:**

The cloud is emerging for scalable and efficient cloud services. To meet the needs of handling massive data and decreasing data migration, the computation infrastructure requires efficient data placement and proper management for cached data. In this paper, we propose an efficient and cost-effective multilevel caching scheme, called MERCURY, as computation infrastructure of the cloud. The idea behind MERCURY is to explore and exploit data similarity and support efficient data placement. To accurately and efficiently capture the data similarity, we leverage a low-complexity locality-sensitive hashing (LSH). In our design, in addition to the problem of space inefficiency, we identify that a conventional LSH scheme also suffers from the problem of homogeneous data placement. To address these two problems, we design a novel multicore-enabled locality-sensitive hashing (MC-LSH) that accurately captures the differentiated similarity across data. The similarity-aware MERCURY, hence, partitions data into the L1 cache, L2 cache, and main memory based on their distinct localities, which help optimize cache utilization and minimize the pollution in the last-level cache. Besides extensive evaluation through simulations, we also implemented MERCURY in a system. Experimental results based on real-world applications and data sets demonstrate the efficiency and efficacy of our proposed schemes.