HIGHLIGHTS

RUI REN, CHUNGHSUAN WU, ZHOUWANG FU, TAO SONG, YANQIANG LIU,
ZHENGWEI QI, AND HAIBING GUAN

In large-scale data-parallel analytics, the shuffle phase is crucial and heavily affecting the end-to-end application performance. To reduce shuffle overhead, we present *SCache*, an open-source plug-in system that particularly focuses on shuffle optimization. We also propose a new performance model called *Framework Resources Quantification* (FRQ) model to analyze DAG computing frameworks and evaluate the SCache shuffle optimization. The main contributions of this manuscript are:

(1) SCache decouples the shuffle write and read from both map and reduce tasks. Such decoupling effectively enables more flexible resource management and better multiplexing between the computational and I/O resources.
(2) SCache pre-schedules the reduce tasks without launching them and pre-fetches the shuffle data. Such pre-scheduling and pre-fetching effectively overlap the network transfer time, desynchronize the network communication, and avoid the extra early allocation of slots.
(3) The FRQ model quantifies computing and I/O resources and visualizes the resources scheduling strategies of DAG frameworks in the time dimension. We use the FRQ model to assist in analyzing the deficiencies of resources scheduling and optimize it.
(4) The performance of SCache has been evaluated with both simulations and testbed experiments on a 50-node Amazon EC2 cluster. Those evaluations have demonstrated that, by incorporating SCache, the shuffle overhead of Spark can be reduced by nearly 89%, and the overall completion time of TPC-DS queries improves 40% on average. On Apache Hadoop MapReduce, SCache optimizes end-to-end Terasort completion time by 15%.

In summary, we propose a new data processing solution for the big data analytics and a new performance model. We believe that the content makes a sufficient contribution to this journal submission.