

Summary of Difference

October 11, 2018

This paper is extension of our conference paper at PPOPP '18 [1]. This paper makes the following distinct contribution compared with the conference paper.

1. FRQ model quantifies computing and I/O resources and displays them in time dimension. Furthermore, FRQ model can calculate execution time of DAG framework under difference resources allocation strategies.
2. In conference paper, we only implement SCache on Spark. To prove SCache compatibility as a cross-framework plug-in, we also implemented SCache on Hadoop MapReduce. By employing shuffle data management, SCache can also optimize the computing performance of Hadoop MapReduce.
3. In experiment, we append two parts. First we verify FRQ model under two different environments, the error between calculated value and experimental value is basically below 10%. Furthermore, we evaluate the performance of SCache on Hadoop Mapreduce in the same environment as conference paper. After utilizing pre-fetching of SCache, Hadoop Mapreduce with SCache optimize overall job execution time by 15%.

Besides the major contributions, we also revise the paper in many places, including writing and more related work. We believe that the added content makes sufficient contribution to this journal submission.

References

- [1] Z. Fu, T. Song, Z. Qi, and H. Guan, “Efficient shuffle management with scache for dag computing frameworks,” in *Proceedings of the 23rd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. ACM, 2018, pp. 305–316.