Abstract

Quantifiability is a concept in MapReduce Analytics based on the following two conditions: (a) a mapper should be cautious, that is, should not exclude any reducer’s shuffle and sort strategy from consideration; and (b) a mapper should respect the reducers’ shuffle and sort preferences, that is, should deem a reducer’s shuffle and sort strategy ki infinitely more likely than k’i if it premises the reducer to prefer ki to k’i. A shuffle and sort strategy is quantifiable if it can optimally be chosen under common shuffle and sort conjecture in the events (a) and (b). In this paper we present an algorithm that for every finite MapReduce operation computes the set of all quantifiable shuffle and sort strategies. The algorithm is based on the new idea of a key-value preference limitation, which is a pair (ki, Vi) consisting of a shuffle and sort strategy ki, and a subset of shuffle and sort strategies Vi, for mapper i. The interpretation is that mapper i prefers some shuffle and sort strategy in Vi to ki. The algorithm proceeds by successively adding key-value preference limitations to the MapReduce.


Wittawat Tantisiriroj, Swapnil Patil, and Garth Gibson. Data-intensive file systems for Internet services: Arose by any other name…. Technical Report CMU-PDL-08-114,Parallel Data
Laboratory, Carnegie Mellon University, 2008.
- Howard Karloff, Siddharth Suri, and Sergei Vassilvitskii. A model of computation for
(SODA 2010), Austin, Texas, 2010.
- Thomas Sandholm and Kevin Lai. MapReduce optimization using regulated dynamic
and Modeling of Computer Systems (SIGMETRICS &apos;09), pages 299-310, Seattle,
- M. Mustafa Rafique, Benjamin Rose, Ali R. Butt, and Dimitrios S. Nikolopoulos.
Supporting MapReduce on large-scale asymmetric multi-core clusters. ACM
- Jeffrey Dean and Sanjay Ghemawat. MapReduce: A flexible data processing tool.
- Jonathan Cohen. Graph twiddling in a MapReduce world. Computing in Science and
- Michael Stonebraker, Daniel Abadi, David J. DeWitt, Sam Madden, Erik Paulson,
Andrew Pavlo, and Alexander Rasin. MapReduce and parallel DBMSs: Friends or foes?

Index Terms

Computer Science

Algorithms

Keysords

MapReduce analytics quantifiability key-value preference limitation shuffle and
sort	Totally
Ordered Data-Intensive Systems