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 $k_i$ infinitely more likely than $k'_i$ if it premises the reducer to prefer $k_i$ 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 $(k_i, V_i)$ consisting of a shuffle and sort strategy $k_i$, and a subset of shuffle and sort strategies $V_i$, for mapper $i$. The interpretation is that mapper $i$ prefers some shuffle and sort strategy in $V_i$ to $k_i$. The algorithm proceeds by successively adding key-value preference limitations to the MapReduce.
Characterization of Randomized Shuffle and Sort Quantifiability in MapReduce Model


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.


**Index Terms**

Computer Science
Algorithms

**Keywords**
MapReduce analytics   quantifiability   key-value     preference limitation   shuffle and sort

Totally Ordered Data-Intensive Systems