اسماء المؤلفين:

RadhyaSahal, Mohammed H. Khafagy and Fatma A. Omara

ملخص البــــــــحث باللغة الانجليزية :

Mostly, sorting of data is a common operation in many applications, which causes the consumption of resourcesand thus leads to computation overheads. Regarding the context of Big Data multi-query, the shared sort operationsare fairly large, which incur high-cost I/Os whether explicit or implicit. In particular, Big Data multi-query,including aggregation and sort operations, takes long execution time due to reshuffle of the samedata multiple times using similar tasks. Therefore, exploiting the sharing data and the sharing sort opportunitiesof similar tasks can offer the possibility of reusing the previous results to optimize multi-query. For consideringsharing data, our previous work, Multi-Query Optimization Using Tuple Size and Histogram (MOTH) system, hasbeen introduced to consider the granularity of the sharing data opportunities among multi-query. However, timeoverheads regarding redundant data in-network movement (i.e., shuffling time to transfer intermediate data forsort operations) have not been considered. Therefore, the MOTH system has been extended to SOOM (Sort-Based Optimizer over MOTH) system to exploit sharing sort opportunities, including explicit sorts of sort queriesand implicit sorts of aggregation queries. The proposed SOOM system consists of two additional modules toexploit sharing sort opportunities, namely query explorer and sort exploiter, which leverage our existingMOTH system to fulfill optimizing multiple aggregation and sort

queries. The experimental evaluation hasshown that the SOOM system outperforms the naive and the state-of-art techniques regarding query execution

time among queries by 45% and 30%, respectively, while introducing maximal intermediate data size reductionby 67% and 61% in average, respectively, over Hadoop-like infrastructures

البحث مشتق من رسالة علمية

يقع البحث ضمن مجالات البحث بالقسم العلمي