عنوان البـــــحث باللغة الانجليزية:

iHOME: Index-Based JOIN Query Optimization for Limited Big Data Storage

اسماء المؤلفين:

RadhyaSahal · MarwahNihad · Mohamed H. Khafagy · Fatma A. Omara

ملخص البـــــحث باللغة الانجليزية :

Query optimization in Big Data becomes a promising research direction due to the popularity of massive data analytical systems such as Hadoop system. The query optimization is getting hard to efficiently execute JOIN queries on top of Hadoop query language, Hive, over limited Big Data storages. According to our previous work, HiveQL Optimization for JOIN query over Multi-session Environment (HOME) system has been introduced over Hadoop system to improve its performance by storing the intermediate results to avoid repeated computations. Time overheads and Big Data storages limitation are considered the main drawback of the HOME system, especially in the case of using additional physical storages or renting extra virtualized storages. In this paper, an index-based system for reusing data called indexing HiveQL Optimization for JOIN over Multi-session Big Data Environment (iHOME) is proposed to overcome HOME overheads by storing only the indexes of the joined rows instead of storing the full intermediate results directly. Moreover, the proposed iHOME system addresses eight cases of JOIN queries which classified into three groups; Similar-to-iHOME, Compute-on-iHOME, and Filter-of-iHOME. According to the experimental results of the iHOME system using TPC-H benchmark, it is found that the execution time of eight JOIN queries using iHOME

on Hive has been reduced. Also, the stored data size in the iHOME system is reduced relative to the HOME system, as well as, the Big Data storage is saved. So, by increasing stored data size, the iHOME system guarantees the space scalability and overcomes the storage limitation

البحث مشتق من رسالة علمية

يقع البحث ضمن مجالات البحث بالقسم العلمي