

JOUM: An Indexing Methodology for Improving Join in Hive Star schema

International Journal of Scientific & Engineering Research, Volume 6, Issue 3, PP 111-119, March-2015

أسماء المشتركين في البحث :

Hussien SH. Abdel Azez- Mohamed H. Khafagy- Fatma A. Omara

ملخص البحث باللغة الانجليزية

Now a day, big data represents an important and complex Issue of information extraction/retrieval due to required analysis computation power. Also, Database Star schema is considered one of the complicated data models due to the using of Join queries to extract information and generate requested reports. These Join queries need to scan a vast amount of data (tera, peta, zet a byt es). On the other hand, HIVE is one of the important and efficient big data SQL querying tools built on the top of Hadoop to translate SQL queries into Map/Reduce tasks. By using indexing data of Join queries could speed up HIVE Join query (map/reduce) tasks especially in Star Schema. According to the work in this paper, JOUM (Join once Use Many) methodology has been introduced to pre-join the star schema data and build an index of Joined data. Based on JOUM, SQL queries execution time in HIVE has been improved without changing HIVE framework. TPC-H benchmark has been used to evaluate the performance of JOUM methodology. The experimental result proves that JOUM methodology outperforms traditional Join execution time. Also, JOUM performance is improved by increasing data size. Generally, JOUM can be considered one of the suitable methodologies for big data analysis.