

## **HOME : HiveQL Optimization in Multi-Session Environment**

**5th European Conference of Computer Science (ECCS '14),  
Pp 80:89, 2014**

أسماء المشتركين في البحث :

**Marwah N. Abdullah- Mohamed H. Khafagy- Fatma A.  
Omara**

ملخص البحث باللغة الانجليزية

Analyzing big data has emerged as a significant activity for many organizations. This big data analysis is simplified by the MapReduce framework and execution environment, such as Hadoop and parallel systems, such as Hive. On the other, most of the MapReduce users have a complex query analysis that has expressed as individual MapReduce jobs. By using high-level query languages such as Pig, Hive, and Jaql, the user complex query expresses into workflows of MapReduce jobs. The work in this paper concerns about how to reuse the previous results in the hive output file in the same or different sessions to improve the Hive performance. This has been done by introducing an algorithm called HOME (HiveQL Optimization in MultiSession Environment). To evaluate our developed HOME algorithm, it has implemented using 19 Different SQL Statement to reduce I/O in MapReduce Job. By developing HOME algorithm, a new HiveQL execution architecture based on materialized previous results has proposed. The framework implementation has built on top of the hive dataflow system without any change in Hive. To evaluate the proposed HiveQL architecture performance, the Star Schema benchmark SSB has been used. According to the experimental results, it is found that the performance of the developed HOME algorithm outperforms the Hive an estimated 67% on average.