Abstract

Data warehouse is a centralized repository for analyzing and storing huge amount of data. In distributed data warehouse, data can be shared across multiple data repositories which belong to one or more organizations. Query sorting is one of the issues for formatting the number of queries that can be selected together. Reducing the usual completion period of a random order is a common concern. In this paper, we are dealing three scheduling algorithms for query scheduling and the performance report based on processing time and memory size is also evaluated. The algorithms discussed are Optimal Resource Constraints (ORC), Grouping based Fine-grained Job Scheduling (GFJS) and Heuristic Algorithm (HA). ORC allocates
queries according to their processor's capabilities. GFJS is based on resource characteristics. HA selects some possible schedules that are having the shortest sum of completion time and this set contains the optimal one.

References

- Yeqing Liao and Quan Liu, "Research on Fine-grained Job Scheduling in Grid
Query Processing in Distributed Data Warehouse using Scheduling Algorithms


Index Terms

Computer Science
Datamining

Keywords

Data Warehouse
Optimal Resource Constraints (orc)
Grouping Based Fine-grained Heuristic
Job Scheduling (gfjs)
Algorithm (ha)