AGENT-BASED APPROACH TO SELECTING VIEWS AND INDEXES IN A DATA WAREHOUSE

G.A. Aderounmu, I.O. Awoyelu and A.O. Adetoye

Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.

ABSTRACT

Selection of views and indexes in a data warehouse is an important issue in designing a data warehouse. Proper selection of views and indexes to materialize in a data warehouse results in minimal search time, minimal processor overhead and reduced maintenance cost. In this paper, we present two models experiment (Reactive and Proactive) for selecting views and indexes in a data warehouse based on Agent Technology. A sample query was defined using the Structured Query Language (SQL). To advance a proof of a superior technique between the two models provided by the agent, a simulation program was developed in Java. Parameters, such as views and index materialization time and processor overhead, were used to measure the performance of the agent technology. Based on the simulation data, mathematical models for search time and processor overhead were derived. From the comparison of the simulation results, it was found that for both models, as the data source size increased, both the search time and the processor overhead increased. Consequently, the drastic reduction in the rate of increase of the views and indexes materialization time for Proactive Agent model makes it a preferred option for selecting views and indexes in a data warehouse. In conclusion, Proactive Agent model is recommended for the implementation of views and indexes selection in a data warehouse, as it is robust, scalable and superior to the Reactive Agent model, while existing schemes failed to achieve practically reduced search time and minimal processor overhead.

Keywords: agent, search time, processor overhead, data warehouse, data source.

INTRODUCTION

The concept of "data warehousing" dates back at least to the mid-1980s, and possibly earlier [1]. In essence, it was intended to provide an architectural model for the flow of data from operational systems to decision support environments. A data warehouse can be defined as a copy of transaction data specifically structured for querying and reporting [2,3]. A data warehouse is defined in [4] as a collection of the pieces of information used to manage and direct the business for the most profitable outcome. In the above definitions, data warehouse is more than just data; it is also the process involved in getting that data from tables to the analyst. As a working definition, in this paper, a data warehouse is defined as an integration of information from a number of data sources that can be used for querying and data analysis by users.

One of the most attractive activities within a data warehouse is the selection of views and indexes. A view is just a named query [5]. A materialized view is a table containing the query result. A data warehouse stores one or more materialized views of the source data [6]. The data is then readily available to user application for querying and analysis of the end users. The existing schemes such as uniform approach, and the greedy approach for selecting views and indexes have their limitations. According to [7], the selection of an appropriate set of views using existing schemes failed to minimize total query response time, cost of maintaining the selected views, and processor overhead practically.

This paper focuses on the development of a new technique for selecting views and indexes in a data warehouse using the concept of agent technology. The main justification for using this technique is its practicability. To our knowledge, there has been no similar direct technique for selecting views and indexes in a data warehouse.

This paper is organized as follows. In the following section, we described the concept of agent technology used for the two models developed in this paper. Also, we reported the existing schemes for selecting views and indexes in a data warehouse and highlight some of the problems associated with those schemes. In section 3, we discussed the proposed models vis-à-vis Reactive model and Proactive model, while section 4 described the simulation, and mathematical models derived from the simulation were discussed in section 5. In section 6, we discussed the results of the simulation for the two models proposed and in section 7, we ended with some conclusions.