An Approach to Specify When Reselecting Views to be Materialized

Henda Ben Ghezala*, Abdelaziz Abdellatif**, Ali Ben Ammar ***

*Ecole Nationale des Sciences de l'Informatique , 2010 Manouba. Tunisie
Henda.bg@cck.rnu.tn
**Faculté des Sciences de Tunis, 2092 Tunis. Tunisie
abdelaziz.abdellatif@fst.rnu.tn
***Institut Supérieur de Documentation, 2010 Manouba. Tunisie
ali.benammar@isd.rnu.tn

Abstract: A data warehouse stores a large volume of data extracted from multiple sources. A set of materialized views is defined over the base tables in order to optimize OLAP (On-Line Analytical Processing) query response time. The selection of materialized views may be static or dynamic. The dynamic selection is continually controlled by a system that calibrates the set of views. The static selection is controlled periodically by the data warehouse administrator who provides the parameter values to the view selection program and defines the selection period. A short period may increase the system workload if there are unnecessary executions of the view selection program. A long period may decrease the query response time. In this paper we propose an algorithm to specify when to select views to be materialized in a static policy. That is when the view selection program should be run. Our main contribution is the use of some tolerance parameters to update and reselect the materialized views. The materialized views will be updated only when it is necessary. The view selection program will be executed either at the end of the selection period, defined by the administrator, or when there is a non tolerated increase of the query execution cost. The aim is to reduce the materialization cost and to guarantee a high query response time. Our experiment results show that, for some values of the tolerance parameters, our approach is more profitable than the static view selection algorithms.

1. Introduction

A data warehouse stores a large volume of data extracted from multiple sources. A set of materialized views is defined over the base tables in order to optimize OLAP (On-Line Analytical Processing) query response time. The selection of these materialized views is based on a cost model that combines, in general, the view maintenance cost and the query execution one. It respects a given limited amount of resources such as materialization time, storage space, or total view maintenance time.

The selection of materialized views may be static [Agrawal et al., 2000][Gupta,1997][Gupta et Mumick, 1999][Gupta et al., 1997][Harinarayan et al., 1996