Efficient Materialized View Selection for Multi-Dimensional Data Cube Models

Naveen Dahiya, Maharaja Surajmal Institute of Technology, New Delhi, India
Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India
Manjeet Singh, Y.M.C.A. University of Science and Technology, Faridabad, India

ABSTRACT

Decision Support Systems help managers to make intelligent decisions by throwing complex queries on large databases. The response time to queries is a very crucial factor in governing the quality of decision support systems. The response time can be greatly improved by using query optimization techniques. A powerful query optimization technique selects only some of the views and not all views for materialization. The authors in this paper present a refined greedy selection approach using forward references to give better materialized view selection. The approach works on lattice framework of data that is capable enough to show inter dependencies of data. The choice of materialized views using the proposed approach gives a better trade off in terms of space/benefits, which is proved from the experimental results. The refined greedy selection approach is independent of space constraint and depends on number of passes entered by the user. The view selection is further enhanced by including space constraints to the results of greedy and refined greedy approach using knapsack implementation.

KEYWORDS

Data Cube, Decision Support Systems, Greedy Selection, Knapsack Approach, Lattice, Materialization, Query Optimization, Views

1. INTRODUCTION

Decision Support Systems (Harinarayan et al. 1996) are used by all business enterprises to gain competitive advantage in current market scenario. All the decision support systems are supported by huge data warehouses at the backend. Complex queries run on data warehouses and results achieved. Query response time is one of the major factors affecting the quality of data warehouses. A number of query optimization (Rafanelli 2003; Parimala and Kumar 2002) techniques exist of which one such technique involves greedy selection of views that defines the order of view selection. The greedy approach (Harinarayan et al. 1996) selects the view that has maximum cost benefit among all the views not selected so far. The greedy approach aims at selecting a fixed number of views to be materialized irrespective of space constraint.

The authors aim to select materialized views following a refined greedy approach. At each pass one view is selected that gives maximum cost benefit. When more than one views with same maximum cost benefits exist, then the choice of views is made by selecting the view that gives maximum cost benefit in next subsequent pass. If multiple views with same maximum cost benefits exist in the subsequent pass, then that view is selected corresponding to which a view has maximum cost benefit and minimum occupied space in the subsequent pass. The proposed approach considers cost benefits along with space occupied in greedy selection of views, makes use of forward referencing and is capable enough to give better results than existing greedy approach. The results achieved using refined greedy approach is then used for view selection under space constraints using knapsack implementation.
1.1. Preliminaries

Few terms are used throughout the content of paper and should be known to understand the paper and its implications in research domain. The terms are as follows:

**View**: A derived relation/result in response to a query. It is defined in terms of base relation and/or combination of attributes (Mami and Bellahsene 2012). Each cell in multidimensional cubes forms a view.

**Materialized View**: A view is materialized if its result in response to query is stored in memory. It may be local copy of remotely located data items, or a result of joins/aggregations (Mami and Bellahsene 2012; Ye et al 2005). It is the set of materialized views whose optimal selection improves query optimization.

**View Selection**: It aims at selecting a set of materialized views given some database to optimize query response time (Mami and Bellahsene 2012). The optimal view selection improves query response time and is one of the main factors affecting query optimization of decision support systems.

**Data Cubes**: The data in a data warehouse is viewed along multiple dimensions. Multidimensional analysis of data is graphically represented as data cubes (Mami and Bellahsene 2012). The data cube is the source for view generation.

**Lattice**: A Lattice (Shukla et al 2000) is a graphical framework used to show dependencies among multiple views of a multi dimensional data warehouse. Lattice is input for materialized view selection approaches. A lattice is a ordered collection of views to which view selection approaches are applied to get optimal subset of materialized views. The optimal selection of views enhances query optimization. Lattice framework is explained in detail in the successive sections of the paper.

1.2. Related Research and Motivation

Query Optimization has been a major concern for researchers in data warehouse domain. One of the major factors affecting query optimization is optimal selection of materialized views (Bellahsene et al 2012). Many solutions to the view selection problem have been proposed and analyzed (Dhote and Ali 2009; Halevy 2001; Harinarayan et al 1996; Kumar and Kumar 2012a; Kumar and Ghoshal 2009). There exist several issues related to materialize view selection. Dhote and Ali 2009 provide analysis of methodologies for materialized view selection in data warehouse systems. The first framework for view selection using greedy approach was proposed by Harinarayan et al 1996. He discuss lattice framework, cost model, benefit metric and greedy approach for materialized view selection which are taken up in detail in subsequent sections in the paper. A comparison is also made for greedy view selection and optimal view selection by Harinarayan et al 1996. A pick aggregates algorithm for view selection based on greedy approach is proposed by Shukla et al 2000. The algorithm selects aggregate of views based on pre computed benefits following greedy approach. Many researchers have used A* algorithm (Nilsson 1971) based approach to materialize view indexes (Ross et al. 1996). The solutions to materialized view selection (Mami and Bellahsene 2012; Kumar and Devi 2012) are categorized along various dimensions like: frameworks and resource constraints. Mami and Bellahsene 2012 have categorized various algorithms employed to perform view selection as: deterministic algorithms, randomized algorithms, hybrid algorithms or constraint programming. These algorithms differ in their approach to solve materialized view selection problem and so differs their time and space complexities.

The greedy selection approach as proposed by Harinarayan et al 1996 motivated the authors to propose a more refined greedy selection approach by including some more parameters like space
Query Expansion Based on Central Tendency and PRF for Monolingual Retrieval

Crosslingual Access to Photo Databases
Christian Fluhr (2012). *Next Generation Search Engines: Advanced Models for Information Retrieval* (pp. 78-91). www.igi-global.com/chapter/crosslingual-access-photo-databases/64421?camid=4v1a