



## Cost Effective Algorithms for Materialized View Selection in Data Centric Environment

Sachin D Choudhari\*  
CSE & RGPV  
Bhopal, India

Mrs Rashi Agrawal\*  
IT & CSJMU  
Bhopal, India

**Abstract**— *The notion of data warehouse can be defined as a central repository for all or significant parts of the data that an enterprise's various business systems collect which support making management's decision effectively, for the success of data warehouse accurate and timely consolidated information as well as quick query response times is the fundamental requirement. To avoid accessing from base table and increase the speed of queries posted to a Data warehouse, we can use some pre-computed intermediate results from the query processing stored in the data warehouse called materialized views. The result of useful Materialized view selection process provides an efficient data warehousing system. However, the materialized view needs to be effectively maintained to keep its contents integrated and consistent with the contents of its data sources. The materialized views are associated with some maintenance cost, that's why materialization of all views is next to impossible.*

**Keywords**— *Data centric Environment, Materialized View, View Preservation, Query access frequency, Threshold.*

### I. INTRODUCTION

One of the most important decisions in designing data warehouse is selecting only those views to materialize which eliminates the overhead associated with expensive joins and aggregations for a large set of important class of queries. A materialized view is pre-computed data stored in a table that transparently allows users to query huge amounts of data much more quickly than they could access from the base table. Database retrieval of the materialized view is just like a cache, which is nothing but the copy of the data that can be retrieved quickly. To select an appropriate set of view is the important target that reduces the entire query response time, however to maintain the selected views is critical but very important aspect of building effective data warehouse.

The process of reflecting changes to a materialized view in response to the changes (insert or update or delete) in the base relation is called as 'View Maintenance' that incurs a 'View Maintenance Cost'. Because of view maintenance cost, it is not possible to make all views materialized under the limited space constraints. This need to select only appropriate set of views to materialize for answering queries, this was denoted as Materialized View Selection (MVS) and incrementally checking the materialized view for its current usage i.e. finding the current access frequencies of the materialized views which is an important factor for the preservation of the selected view is denoted as Materialized View Preservation (MVP). Materialized views are very important for improving performance in many business applications that's why recently database research community paying attention to the materialized view selection preservation and maintenance process. This paper is organized as follows. We describe a related work of materialized view selection and materialized view preservation in section II, proposed Materialized Views Selection & preservation framework details are explaining in section III. In section IV we concluded the paper and next section is used to list the references.

### II. RELATED WORK

The materialized view selection problem has received significant attention in this literature.

Related researches differ in several points:

- The approach of the set of candidate views is determined;
- The framework used to capture relationships between different candidates views;
- The use of mathematical cost models vs. calls to the systems query optimizer;
- View selection in the relational or multidimensional context;
- Multiple or simple query optimization;
- Theoretical or technical solutions for materialized view creation, preservation and maintenance.

The problem of finding prominent views to materialize to answer frequent queries has been studied under the name of Materialized view selection (MVS).

Dr. T.Nalini et al. proposes a mining algorithm for the selection of materialized views so that query evaluation costs can be optimized as well as maintenance and storage was addressed in this piece of work [1].

Ashadevi and Balasubramanian proposes structure for selecting views to materialize (i.e., View selection problem), which takes in to account all the cost metrics associated with the materialized views selection, including query processing

frequencies, update frequencies, base relation, query access costs, materialized view maintenance costs and the system's storage space constraints and then selects the most cost effective views to materialize and thus optimizes the maintenance storage, and query processing cost. [2]

Himanshu Gupta and Inderpal SinghMumick developed an algorithm to include the maintenance cost and storage constraint in the selection of materialized views for data warehouse environment [3]

Yang, J et al. proposed a heuristics algorithm based on individual most favorable query plans. Framework is based on specification of multiple views processing plan (MVPP), which is used to present the problem formally [4]

Harinarayan et al. proposed an algorithm for the materialized views selection so that query costs estimate can be optimized in the unique case of "data cubes". This paper provides fine combination between the space used and the average time to answer complex query. In this piece of work the costs for view preservation, maintenance and storage were not addressed at all. [5]

Amit Shukla et al. proposed a very simple and fast heuristic algorithm, PBS, to select aggregates queries for pre computation. PBS algorithm runs faster than BPUS, and is fast enough to make the searching of the time-space trade-off feasible during system configuration [6]

Y. D. Choudhary et al. proposes an approach of grouping or clustering the similar queries depending on certain parameters like access frequency to get the result from MV. The proposed work explores the area of query clustering for the selection of materialized view to decrease the query response time and storage space. They propose a novel framework is developed for the selection of MV using query clustering. However, this algorithm did not consider the existing materialized view preservation constraints [10]

Wang, X et al. View maintenance techniques are classified into four major categories: self maintainable recomputation, not self-maintainable recomputation, self maintainable incremental maintenance and not self maintainable incremental maintenance. Self-maintainable Incremental maintenance performs the best in terms of both space usage and number of rows accessed. [17]

### **III. PROPOSED METHODOLOGY FOR MATERIALIZED VIEW SELECTION AND PRESERVATION**

This section elaborates proposed framework approach for the selection of materialized view.

Materialized views are beneficial for the users to quickly get the search results for frequent queries. The ultimate aim behind the proposed materialized view selection framework is to materialize the user views by taking into consideration of query frequency, query processing cost and storage requirement of query.

The first phase of materialized view selection is generation of massive set of records for the data warehouse by using data insertion generator. After generation of massive set of records the second phase is to create all possible set of complex queries for the above set of records. The third phase is to select most prominent queries using WMVS (Weighted Materialized View Selection) algorithm. The fourth and final phase is to preserve most deserving materialized views for fast query processing using PMV (Preservation of Materialized View) algorithm.

The benefit of the proposed algorithm is that it can mine the frequent queries with less computation time. So, we have applied proposed algorithm to user's bunch of queries for finding the frequent queries and their corresponding support value. Then, for all the queries, we maintain a table,  $T$  that contains the frequency obtained from the given algorithm, the query processing cost and spatial cost required. Using this table, then, for all the queries, we maintain a table, containing selection cost  $SQ$  of every query  $Q$  is computed by combining the above three values. The main objective is that the storage cost and query processing cost should be minimized but, the frequency-based cost should be maximized. The reason behind is that, if the query is to be materialized, then the query should be frequently used by the number of users. On the other hand, the storage cost should be minimum in order to reduce the space require to store the results. By considering this multi-objective, given approach is used to find the selection cost and then by using selection cost and user specified threshold value materialized views are created.

#### **A. WMVS Algorithm Steps**

The processing steps of Materialized View Selection are as follows:

Step1: Creation of random set of records for given tables in database through random record generator.

Step2: Creation of all possible set of queries on above created database.

Step3: Evaluate access frequency, storage area and processing time of bunch of input queries.

Step4: Calculate access frequency cost, storage area cost and processing time cost of bunch of input queries.

Step5: Calculate query selection cost of above set of queries according to their access frequency cost, storage area cost and processing time cost along with weighted constants  $FC$ ,  $SC$  and  $PC$ .

Step6: Calculate the threshold value ( $T$ ) of the above process queries by summation of each query selection cost divided by number of queries process.

Step7: Create materialized view of only those queries having query selection cost is greater than threshold value.

#### **B. PMV Algorithm Steps**

The processing steps of Materialized View Preservation are as follows:

Step1: Find access frequency and storage for bunch of existing materialized view.

Step2: Calculate access frequency cost, storage area cost of bunch of existing materialized view.

Step3: Calculate materialized view selection cost of above set of existing materialized view according to their access frequency cost, storage area cost weighted constants  $FC$ ,  $SC$

Step4: Calculate the materialized view threshold value (MT) of the above process queries by summation of each query selection cost divided by number of queries process.

Step5: Delete existing materialized view having materialized view selection cost is lesser than threshold value.

The above proposed algorithm satisfies the materialized view selection constraints very effectively and provides best summary views for optimize query accessing purpose.

#### IV. CONCLUSIONS

As materialized views are used to store the summary precomputed data it is used to improve query performance by minimizing query processing time. On the other side materialized view has associated with some maintenance cost and thus it is impossible to create materialized view of all the user queries. So how to select the set of user queries to be materialized that improve query performance significantly and storage cost for storing materialized view also minimized.

This paper gives the idea regarding how to select a set of materialized view with the help of various significant parameters like: frequency of query, cost of query processing and storage space. Our proposed methodology which determines what types of queries are more beneficial for the creation of materialized view so as to achieve the high query performance and also provide best ways for preservation of created materialized view.

#### REFERENCES

- [1] Dr.T.Nalini,Dr.A.Kumaravel , Dr.K.Rangarajan,” A Novel Algorithm with IM-LSI Index For Incremental Maintenance of Materialized View” JCS&T Vol. 12 No. 1 April 2012
- [2] B.Ashadevi, R.Balasubramanian,” Cost Effective Approach for Materialized Views Selection in Data Warehousing Environment”, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.10, October 2008
- [3] Gupta, H. & Mumick, I., Selection of Views to Materialize in a Data Warehouse. IEEE Transactions on Knowledge and Data Engineering, 17(1), 24-43, 2005.
- [4] Yang, J., Karlapalem. K., and Li. Q. (1997). A framework for designing materialized views in a data warehousing environment. Proceedings of the Seventieth IEEE International Conference on Distributed Computing systems, USA, pp:458.
- [5] V. Harinarayan, A. Rajaraman, and J. Ullman. “Implementing data cubes efficiently”. Proceedings of ACM SIGMOD 1996 International Conference on Management of Data, Montreal, Canada, pages 205--216, 1996.
- [6] A. Shukla, P. Deshpande, and J. F. Naughton, “Materialized view selection for the multidimensional datasets,” in Proc. 24th Int. Conf. Very Large Data Bases, 1998, pp. 488–499.
- [7] Mr. P. P. Karde, Dr. V. M. Thakare. “Selection & Maintenance of Materialized View and It’s Application for Fast Query Processing: A Survey”. Proceedings of International Journal of Computer Science & Engineering Survey (IJCSSES) Vol.1, No.2, November 2010
- [8] Abdulaziz S. Almazayad, Mohammad Khubeb Siddiqui. “Incremental View Maintenance: An Algorithmic Approach”. Proceedings of International Journal of Electrical & Computer Sciences IJECS-IJENS Vol: 10 No: 03
- [9] Y.D. Choudhari and Dr. S. K. Shrivastava, “Cluster Based Approach for Selection of Materialized Views”, International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 2, Issue 7, July 2012
- [10] C. Zhang, X. Yao, and J. Yang.,” An evolutionary Approach to Materialized View Selection in a Data Warehouse Environment.” IEEE Transactions on Systems, Man and Cybernetics, vol. 31, no.3, pp. 282–293, 2001.
- [11] K. Aouiche, P. Jouve, and J. Darmont. Clustering-based materialized view selection in data warehouses. In ADBIS’06, volume 4152 of LNCS, pages 81–95, 2006
- [12] Y. Zhuge, H. Garcia-Molina, J. Hammer, and J. Widom, "View Maintenance in a Warehousing Environment." In Proceedings of the ACM SIGMOD Conference, San Jose California, May 1995.
- [13] Gupta H. Selection of Views to Materialize in a Data Warehouse, in: 6th Int. Conf. on Database Theory (ICDT’97, Delphi, Greece, Jan 8-10), 1997
- [14] Lehner, W., Ruf, T., Teschke, M.: Improving Query Response Time in Scientific Databases Using Data Aggregation, in: Proc. of the 7th Int. Conf. and Workshop on Database and Expert Systems Applications (DEXA’96, Zürich, Schweiz, 9-13 September), 1996
- [15] Abdulaziz S. Almazayad, Mohammad Khubeb Siddiqui,“ Incremental View Maintenance :An Algorithmic Approach”, International Journal of Electrical & Computer Sciences IJECS-IJENS Vol: 10 No: 03
- [16] Jin-Hyuk Yang , and In-Jeong Chung,”ASVMRT: Materialized View Selection Algorithm in Data warehouse”, International Journal of Information Processing Systems, Vol.2, No.2, June 2006
- [17] Wang, X., Gruenwalda. L., and Zhu.G. (2004). A performance analysis of view maintenance techniques for data warehouses. Data warehouse knowledge, pp:1-41.