International Journal of Multidisciplinary Approach and Studies

ISSN NO:: 2348 – 537X

Novel Algorithms for Materialized View Selection and Preservation in Data Warehousing Environment

Mr.Gaurav P. Ambatkar*, Prof. Pratyoosh Raiand Dr. Tripti Arjariya*****

** M. Tech Scholar, Department of Computer Science, Bhabha Engineering Research Institute, RGPV, Bhopal.*

***HOD & Guide, Deptt. of Computer Science & Engineering, Bhabha Engineering Research Institute, Bhopal.*

****Professor, Deptt. of Computer Science & Engineering, Bhabha Engineering Research Institute, Bhopal.*

ABSTRACT-

Data warehousing and online analytical processing represents some of the most up-to-date trends in computing environments to large scale processing and analysis of data. The theory attempted to address the range of problems associated with this flow, mainly the high costs associated with it. In the absence of a data warehousing architecture, an enormous amount of redundancy was required to maintain multiple decision support environments. In larger corporations it was typical for several decision support environments to control independently. Though each environment served different users, they often essential much of the same stored data. The process of gathering, cleaning and integrate data from various source, generally from long-term accessible operational systems. Data warehousing technology is becoming necessary for the efficient business policy formulation and execution. For the success of any data warehouse accurate and timely consolidate information along with immediate and efficient query response times is the basic fundamental constraint.The materialization of all views is nearly impossible because of the materialized view storage space and maintenance cost restriction thus proper materialized views selection is one of the intelligent decision in designing a data warehouse to get best possible efficiency. In this paper, we represent a structure for selecting best materialized view so as to achieve the effective combination of good response time, low processing cost and low maintenance cost in a particular storage space restriction. The framework implementation parameters include frequency cost, storage cost and processing cost. The structure select the best cost successful *materialize views to optimize the query processing time thereby ensuing efficient data warehousing system.*

Keywords- Data Warehouse Materialization, View-Maintenance, Access frequency, Threshold value, Query processing cost.

1. INTRODUCTION

A fundamental requirement for the success of a data warehouse is the ability to make available decision makers with both accurate and timely consolidated information as well as fast query response times. The method that is used in practice for providing higher information and best response time is the concept of quickly answered materialized views.

One of the most important decisions in designing data Warehouse is selecting views to materialize for the purpose of efficiently supporting the decision making. The view selection problem defined is to select a set of derived views to materialize that minimizes the sum of total query response time & maintenance of the selected views. So the goal is to select an appropriate set of views that minimizes total query response time and also maintains the selected views. The decision "what is the best set of views to materialize?" must be made on the basis of the system workload, which is a sequence of queries and updates that reflects the typical load on the system. One simple criterion would be to select a set of materialized view that minimizes the overall execution time of the workload of queries.

A view is defined as a function from a set of base tables to a derived table and the function is recomputed every time the view is referenced. On the other hand, a materialized view is like a cache *i.e.* a copy of data that can be accessed quickly. Utilizing materialized views that incorporate not just traditional simple operators but also complex online analytical processing operators play. Materialized views are helpful in applications such as data warehousing, replication servers, data recording systems, data visualization and mobile systems. In certain situation, it is more profitable to materialize a view than to compute the base tables every time the view is queried. Materializing a view causes it to be refreshed every time a change is made to the base tables that it references. It can be costly to rematerialize the view each time a change is made to the base tables that might affect it. So it is desirable to propagate the changes incrementally *i.e.* the materialized view should be refreshed for incremental changes to the base tables. In the last few years, several view maintenance methods have been designed and developed to obtain an efficient incremental view maintenance plan. This paper presents materialized view selection, in which views are selected at the time of query processing for faster query performance. In next section various recent past work that has been carried out in the field of materialized view selection and their utilization for the query processing are stated.

A data warehouse is a repository of subjectively selected operational data which may successfully answer any complex, statistical or analytical queries. Data warehousing enables easy organization and maintenance of large data in addition to fast retrieval and analysis in the desired manner and depth required from time to time. As the data size increases continuously, the speed requirements for processing the data so as to comprehend The pre-computed intermediate results obtained in the query processing are stored in the data warehouse called as materialize views, to provide effective solution for the queries posted to the data warehouse, which in turn can prevent the users from accessing the base data tables Materialized views required physical storage space acts just like a cache, which is copy of the data that can be retrieved quickly. At the same time, the use of materialized views requires additional storage space and overhead of view maintenance when refreshing the data warehouse. Data warehouse is capable of answering queries and performing analysis in an efficient and quick manner, in the view of the fact that integrated information is directly available at the warehouse with differences already resolved. Though the data warehouse research community provides effective solutions for the problem of representing data in a form suitable for analytical queries.

The primary intent of this research is to develop a framework for selecting views to materialize so as to achieve finer query response in low time by reducing the total cost associated with the materialized views. The projected framework exploits all the cost metrics coupled with materialized views such as query execution frequency, access cost, base-relation update frequency, view maintenance cost and the system’s storage space constraints. The queries with high frequencies are selected for the view selection problem. This paper is organized as in section 2 explain related work of materialized view selection and materialized view maintenance and in section 3 explain Materialized Views Selection framework implementation and in section 4 explain algorithm.

2. RELATED WORK

Different approaches have been used by different researches for improving performance in many business applications that's why recently database research community paying attention to the materialized view selection and maintenance.

Yang, J proposed a heuristics algorithm based on individual optimum query strategy. Structure is based on specification of multiple views processing plan, which is used to present the problem formally.

Dr. T.Nalini proposed an IM-LXI index for incremental maintenance of materialized view selection of materialized views. In which query evaluation costs can be optimized as well as view maintenance and view storage was addressed in this piece of work.

Ashadevi, B and Balasubramanian proposed structure for selecting views to materialize which takes in to account all the cost metrics associated with the materialized views selection, including processing frequencies, base relation, update frequencies, query access costs, 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. This piece of work also includes the preservation of existing materialized view.

Himanshu Gupta and InderpalSinghMumick developed a greedy algorithm to minimize the maintenance cost and storage constraint in the selection of materialized views for data warehouse. In this paper view selection under disk space & maintenance cost constraints are addressed.

Harinarayan proposed a greedy algorithm for the materialized views selection so that query evaluation costs can be optimized in the special case of "data cubes". Also provides good trade-offs between the space used and the average time to answer query. But the costs for view maintenance and storage were not addressed in this piece of work.

P. P. Karde This paper gives an overview of various techniques that are implemented in past recent for selection of materialized view and issues related to maintaining the materialized view are also discussed in this paper. Here somefuture aspects are also stated that might be useful for recent researchers.

Y.D.Choudhari proposed a novel CBFSMV algorithm is proposed for selection of materialized view using query clustering strategy that reduces the execution time as compared to response time for actual database.

3. MATERIALIZED VIEWS SELECTION FRAMEWORK

The first algorithm is used to find candidate queries frequency, processing time and storage space information. In the first step how much storage space is occupied by the materialized view will be calculated and if calculated materialized view storage space value is larger than the allotted materialized view storage value then in that case materialize view creation process discarded due to storage space constraints, or else for each query in query set find the frequency, processing time and storage space that can be stored in query information list in the form of query information .

The second algorithm is used to find the best queries that require to materialize to optimize the complex query processing time. Here the first algorithm output i.e. SQ_{LIST} is used as an input to calculate the query frequency, query processing and query storage cost. Using weighted combination of query frequency, query processing and query storage cost, materialized view selection cost will be calculated for each query. After calculating each query selection cost the next step is to find the appropriate materialized view selection threshold value using summation of all the selection cost divided by number of selected queries. If the selection cost is greater than the materialized view selection threshold value then calculate materialized view storage space and add the selected query storage space to get the total materialized view storage space and if total materialized view storage space is less than the materialized view storage threshold value then build the materialized view for the selected query otherwise neglect the query.

Assumptions: SQ_{set} -Given set of queries, SQ_{af} -Query access frequency, SQ_{sp} -Query storage space, SQ_{pt} -Query processing time, SQ_{il} -Query information list, MV_{sp} -Materialized View storage space, SQ_{id} -Query information DTO, T_S –MV storage threshold value.

3.1. Algorithm1: To find candidate queries frequency, processing time and storage space information.

1.begin:

2. Calculate MV_{SP}

3. if ($MV_{SP} < T_S$) then

Repeat for I ←1 to SQ_{set}

SQ_{DTO}←find SQaf

SQ_{DTO}←find SQsp

SQ_{DTO}← find SQpt

SQ_{LIST}←SQ_{DTO}

end repeat

end if

4 else

Discard the materialize view creation process.

end else

5 end

Assumptions:

T- MV selection threshold value, SQ_{mfreq} -Maximum query frequency, SQ_S -Selected query storage, SQ_{mpt}-Maximum query processing time, SQ_{ms} - Maximum Query storage, SQ_{pc} - Query processing cost, SQ_{sc}- Query storage cost, SQ_{fc}- Query frequency cost, SQ_{ct}- Query cost table, SQc - Query selection cost, MV_{sp} - materialized view storage space, M_T - Minimum threshold ,N -Number of rows in query cost table, N_{QIC}-Number of rows in query information list,w1, w2 & w3 - Weighted constant values in between 0 to 1

3.2 Algorithm 2: To find the best queries that need to materialize to optimize the complex query processing time.

begin:

Repeat for I ← 0 to SQ_{LIST} -1

SQ_{DTO}← SQ_{LIST} [i]

SQ_{af}← SQ_{DTO}

SQpt← SQ_{DTO}

SQsp← SQ_{DTO}

SQfc← *SQaf* /SQ_{mfreq}

$SQpc \leftarrow SQpt / SQ_{mpt}$

$SQsc \leftarrow SQsp / SQ_{ms}$

$SQct \leftarrow SQfc$

$SQct \leftarrow SQpc$

$SQct \leftarrow SQsc$

end repeat

[For Finding selection cost]

Repeat for I $\leftarrow$1 to SQ_{ct}

$SQc = w1 * SQ_{fc} + w2 * SQ_{pc} + w3\ (1 - SQ_{sc})$

$SQ_{ct} \leftarrow SQc$

end repeat

[For Select MV Selection Threshold]

$M_T = \sum^{K}_{i=K} SQc / N$

[For Select materialized view having good query response, low processing and storage cost]

Repeat for i $\leftarrow$0 to N-1

$SQc \leftarrow Q_{ct}[i]$

if ($SQc >= M_T$) then

Calculate MV_{SP}

$MV_{sp} \leftarrow MV_{sp} + SQc$

if($MV_{sp} < T_S$) then

Build the materialized view for the selected query

else

Discard the query

else

Discard the query

end repeat

4. CONCLUSION

The materialized view is useful for improving query performance to stores pre-computed data. The selection of views to materialize is the important issues due to the view maintenance cost. In this Paper we have defined the idea regarding best view selection for materialization and determines which queries are more beneficial using combination of query frequency, processing and storage cost for the creation of materialized view so as to achieve the quick query processing time We have presented proposed method for selecting views to materialize so as to achieve the best combination good query performance.These algorithms are found efficient as compared to other materialized view selection and maintenance strategies.

REFERENCES

I. 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.

II. 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.

III. 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.

IV. 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.

V. 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.

VI. 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.

VII. Wang, X., Gruenwalda. L., and Zhu.G. (2004). A performance analysis of view maintenance techniques for data warehouses. Data warehouse knowledge, pp:1-41.

VIII. 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 (IJCSES) Vol.1, No.2, November 2010.

IX. Abdulaziz S. Almazyad, Mohammad KhubebSiddiqui. “Incremental View Maintenance: An Algorithmic Approach”. Proceedings of International Journal of Electrical & Computer Sciences IJECS-IJENS Vol: 10 No: 03.

X. Elena Baralis, Tania Cerquitelli, and Silvia Chiusano,” I-Mine: Index Support for Item Set Mining” IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 4, April 2009 .

XI. 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.

XII. 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.

XIII. 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.

XIV. 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.