International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Over View of Past Literature and Making Comparative Analysis Along with Our Strategy of Materialized View Selection

Sapna Choudhary[1], Roshna Mahajan[2]

[1]Assistant Professor at Shri Ram Group of Institutions, Jabalpur, Affiliated to RGPV University, Bhopal, India

[2]Student of M.Tech- 4th semester at Shri Ram Group of Institutions, Jabalpur, Affiliated to RGPV University, Bhopal, India

Abstract: ***This study is carried out with the aims of undertaking survey of the different approaches used by the researchers to select materialized view in optimal manner of data warehouse (DW). We have done comparative analysis of the available literature on the basis of relevant evaluation parameters viz. processing time, query frequency and spatial cost, area, storage etc. Our designed approach of materialized view selection is also given in this paper. This study may be helpful to the researchers, who are working in the domain of the data warehouse (DW) focusing on the materialized view selection.***

Keywords: Data warehouse (DW), materialized view, comparative analysis, processing time, query frequency and spatial cost, area, storage

1. Introduction

As we all aware the present markets are much more struggling than ever. Any kind of business either faster leads or even fails as per the complexity or easy way of flow/speed of their, with ability to analyze and synthesize the information by using information systems they adopted. A DW is a subject-oriented, integrated, time-varying, nonvolatile collection of data that is used primarily in organization decision making [1]. As an emerging network service, a data warehouse system collects data from many data sources through communication networks locally and internationally by adopting an update-driven approach. By using DW system we can consolidate historical data for analysis and access such analysis by users to use locally and remotely.

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 understand the meaning of this data are also required to be increase significantly.

When a view is defined, the database system stores the desired view itself, and not the result of evaluation of the relational algebra expression that defines the view. Hence, one can define view; view is a derived relation from base relations. A view thus generates a function (derived table) from a set of base tables; this function is typically recomputed every time the view is referenced.

According to the definition of materialized views, the contents of the data warehouse are treated as a set of materialized views defined over the data sources. One can designed the materialized views based on the user's requirements. The benefit of using materialized views is significant since database access to the materialized view is just a cache, which is copy of the data that can be accessed quickly. Other benefits from materialized views are Integrity checking and query optimization. In short, database store the result of query defining the view. Against this, a materialized view is a view whose contents is pre-computed, stored and thus it is optimal in many cases to access the contents of a materialized view than to compute the contents of the view by executing the query defining the view.

As it is impractical to maintain materialized views for all queries due to the huge disk-space consumption and/or large update cost, the DW designer/administrator is facing the problem of selecting view to materialize in the DW. Thus, the issue, constraint and challenges lies in-front of designer is how to select such a set of materialized views with optimal condition of parameters like low processing time, low storage area/space, and having high frequency as the materialized view creation and selection is based on the parameters defined etc.

2. Earlier Related Work

1) Jin-Hyuk Yang et. al. [2], has proposed algorithm – ASVMRT (Algorithm for Selection of Views to Materialize using Reduced Table). From the dimension of given table, firstly, it finds high density clusters then after it produces the reduced tables from the found clusters. Successively, the MVPP is produced from the reduced table and lastly, materialized views are selected from the MVPP in accordance with cost estimation. In this way we can say proposal of paper of ASVMRT adopts clustering method, one of the data mining techniques

2) Algorithms are developed by Himanshu Gupta and Inderpal Singh Mumick [3] in order to select a set of views that can be materialize in a data warehouse to minimize the total query response time with constraint of a given total view maintenance time. They explains an incremental view selection approach address the problem of making a view set self-maintainable by adding auxiliary views. All these problems fall under the general problem that we call the

general view selection problem. Hence, the general view selection problem is the problem of selecting views for materialization in order to satisfy a number of design goals.

3) To achieve the best combination of good query response, low query processing cost and low view maintenance cost, Ashadevi, B and Balasubramanian [4] proposed a framework for selecting views to materialize as it is not possible to materialize all view. But again it has a constraint of given storage space. The framework for selecting views to materialize (i.e. View selection problem), takes in to account all the cost metrics associated with the materialized views selection, including query 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.

4) Elena Baralis et. al.[5], This paper presents the IMine index, a general and compact structure which provides tight integration of item set extraction in a relational DBMS. To reduce the I/O cost, data accessed together during the same extraction phase are clustered on the same disk block. In particular, IMine data access methods currently support the FP-growth and LCM v.2 algorithms. The IMine index has been integrated into the PostgreSQL DBMS and exploits its physical level access methods. Experiments, run for both sparse and dense data distributions, show the efficiency of the proposed index and its linear scalability also for large datasets.

5) Qingzhou Zhang, Xia Sun, Ziqiang Wang [6] shows in his paper materialized view selection algorithm which is based on memtic algorithm (MA). Experimental results show that the proposed algorithm achieves much better performance than other related algorithms.

6) karde Mr. P. P. Karde et al. [7] discuss thoroughly the selection and maintenance of materialized view. They stated that a data warehouse uses multiple materialized views to efficiently process a given set of queries but materialized views selection is one of the crucial decisions in designing a data warehouse for optimal efficiency. So selecting a suitable set of views that minimizes the total cost associated with the materialized views is the key component in data warehousing.

7) Ravindra N. Jogekar & Ashish Mohod [8] present a papers showing a framework for selecting best materialized view so as to achieve the effective combination of good query response time, low query processing cost and low view maintenance cost in a specified storage space constraint. The framework implementation parameter includes query frequency cost, query storage cost and query processing cost. The framework select the best cost effective materialize views to optimize the query processing time thereby resulting efficient data warehousing system.

8) Ashish Mohod et al. [9] shows the survey regarding improve query Performance using effective materialized view selection and maintenance. As per paper the query frequencies, query space, query processing time are the constraints that are the most important factors while selecting rhe views to be materialized

9) For grouping or clustering the similar types of queries, Y. D. Choudhary et al. [10] proposes an approach. This approach depends upon the parameters like access frequency to find the result from MV. His idea look at the area of query clustering for the selection of materialized view with the aim to decrease the time taken and required less storage space. How ever no any idea given for preservation of MV by the approach.

10) In order to achieve the optimal characteristics viz. low storage cost, low query processing cost and high frequency of query and updation of materialized view using LSI, Dr.T.Nalini et al. [11] addressed view selection problem and materialized view maintenance problem by means of taking into account the essential constraints for selecting views to materialize. It removed the constraint of maintenance of materialized view (MMV) also along with selection of it.

11) Dr.T.Nalini et al. [12] propose a model for optimal selection cost i.e minimum cost by taking in to account low storage cost, low query processing cost and high frequency of query by adopting efficient mine algorithms for materialized views selection in a Data Warehouse Environment

3. Our Strategy of Materialized View Selection

In recent times, several algorithms have been proposed by researchers for keeping the views up-to-date in response to the changes in the source data. All have a sort of limitation/constraints. Therefore, we present an improved algorithm for MVS. In our approach, we have designed a mathematical model to select materialized view by considering the frequency, processing cost and area cost. By considering multi-objective, given designed formula is used to find the selection cost (S.C):

$$\mathbf{S.C=W_1xF.C.+W_2x(1-A.C)+W_3x\{cos(PC/60924)-P.C.\}}$$

Where F.C is the frequency cost that can be calculated as frequency of particular query/ max frequency from all queries

Area is calculated as multiplication of row and column of query and A.C is area cost that can be calculated as area to be selected of particular query/ maximum area from all queries P.C is the processing cost that can be calculated as processing time of particular query/maximum processing time from all queries Where, W_1, W_2 and W_3 are Weights such that their sum is equals to 1.Moreover, F.C, A.C *and* P.C represents query frequency cost, query storage area cost and query processing cost. Then, the set of queries that are satisfied the minimum threshold (*T.M.*) is selected to build the materialized.

The threshold value T.M. is calculated by the sum of S.C of all query/total number of query. The required condition is S.C is greater than or equal to T.M, for the view to be materialized

4. Comparative Study

Up to, we have analyzed the various research works; this section gives the comparison of approaches designed by different researchers on materialized view selection for ready reference for the researchers working on materialized view of selected views of data warehouse.

Yang & chung (2006)	View selection	Attribute value density + clustered tables + Selection of views based on clustered/reduced tables	ASVMRT algorithm for view selection	Faster computation time + reduced storage space + 1.8 Times performance better than conventional algorithms	Maintenance of reduced tables not addressed + Updating reduced tables needs attenuation
Gupta & Mumick (2005)	View selection	View selection under disk space & maintenance cost constraints	AND/OR View Graphs + Greedy heuristics Based algorithms	Optimal solution for special cases(AND/OR views) + Polynomial time heuristics	Approximation in view selection Problem not addressed +problem in AND view graphs not NP-hard + Solution fairly Close to optimum
Ashadevi & Balasubramanian (2008)	View selection	Cost effective view selection under storage space constraints	Framework for selecting views + Algorithm For the same + Cost metrics	All cost metrics considered	Query response time not considered + Threshold value not indicated clearly
Elena barlis, Tania cerguitelli,and silvia chiusano (2009)	View selection	Cost effective view selection under storage space constraints	i-mine algorithm for selecting views	Faster Computation time	More memory space
Qingzhou zhang & xia sun,zigiang wang(2009)	View selection	Cost effective view selection under storage space constraints+	MA algorithm for selecting views	Faster computation time + Comparison of Ga & HA algorithm	Only optimal research
Karde & thakre (2010)	View selection	Query cost, maintenance cost, Storage Space	Algorithm for creation and maintenance of views + Algorithm for node selection	Query Performance improved	Only distributed environments highlighted
T.Nalini & A.Kumaravel (2011)	View selection		Cost effective view selection under storage space constraints	i-mine algorithm(modification) for selecting views +using multiple Constraints to reduce storage space	Faster computation time + Reduced Storage space
Dr.T.nalni, Dr.A.Kumaravel(2012)	View selection & maintenance	Cost effective view Selection based on best combination of low storage cost ,low query processing cost and high frequency of query + Updation of materialized view using LSI(Latent Semantic Index)	IM-LSI(Item set mining using Latent Semantic index) Algorithm	. Faster computation time + Reduced Storage space	Selection of threshold value is not calculated
Dr.Y.D. choudari. Dr.S.K. Shrivastava (2012)	View selection	Cost effective view selection under storage space constraints	CBFSMV Algorithm for selection of view	Faster computation time + Reduced storage space	View maintenance problem not addressed

5. Conclusion

In general intension of all researchers lies on how to optimize the materialized view selection. Some papers also focus on the maintenance of selected materialized views. The optimized condition arrived only with have high frequency, low processing time and low storage space- area taken by selected view to materialized it. All papers have sorts of limitation that can be the scope for further research in order to remove the constraints. We have also presented our designed approach in order to select materialized view.

References

[1] Inmon, W.H., 1996. Building the Data Warehouse. 2nd Edn., John Wiley and Sons, Canada, ISBN: 0471-14161-5
[2] Jin-Hyuk Yang, and In-Jeong Chung, June 2006, 'ASVMRT: Materialized View Selection Algorithm in Data Warehouse', International Journal of Information Processing Systems, Vol.2, No.2
[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] 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
[5] Elena Baralis, Tania Cerquitelli, Silvia Chiusano, "IMine: Index Support for Item Set Mining", *IEEE Transactions on Knowledge & Data Engineering*, vol.21, no. 4, pp. 493-506, April 2009, doi:10.1109/TKDE.2008.180
[6] Qingzhou Zhang , Xia Sun & Ziqiang Wang , "An Efficient MA-Based Materialized Views Selection Algorithm" IITA International Conference on Control, Automation and Systems Engineering, CASE 2009, pp: 315-318
[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 (IJCSES) Vol.1, No.2, November 2010
[8] Ravindra N. Jogekar & Ashish Mohod: "Design and Implementation of Algorithms for Materialized View Selection and Maintenance in Data Warehousing Environment" International Journal of Emerging Technology and Advanced Engineering , Volume 3, Issue 9, September 2013
[9] Ashish Mohod & Manoj Chaudhari: "Improve Query Performance Using Effective Materialized view Selection and Maintenance: Survey" International Journal of Computer Science and mobile Computing, Vol.2, Issue4, April 2013 pg.485-490
[10] 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
[11] 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
[12] Dr.T.Nalini, Dr.A.Kumaravel, Dr.K.Rangarajan, "An Efficient I-MINE Algorithm for Materialized Views in a Data Warehouse Environment", IJCSI, Vol.8, Issue5, No1, September 2011

Author Profile

Mrs. Sapna Choudhary received the Bachelor degree and Master degree in Computer Science and Engineering department in 2004, & 2008 respectively from RGPV University, Bhopal, M.P. India. She is currently working as Assistant Professor in department of computer Science and Engineering at Shri Ram Group of Institutions, Jabalpur, affiliated to RGPV University, Bhopal, India

Ms Roshna Mahajan received the Bachelor degree in Computer Science and Engineering department from H.C.E.T, college, Jabalpur, RGPV University, Bhopal, M.P. India in 2006. I am presently doing my master degree in the same Engineering department from S.R.G.I, college, Jabalpur in the same university. During the period of 2006-2011, I have a teaching experience in different polytechnic college of M.P. state