Hema S. Botre, Prof.M.S.Chaudhari / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 3, May-Jun 2013, pp.950-952

Design of an Algorithm For Materialized View Selection

Hema S. Botre
Smt.Bhagwati Chaturvedi
College of Engg,Nagpur

Prof.M.S.Chaudhari
Smt.Bhagwati Chaturvedi
College of Engg,Nagpur

Abstract:

Quick responds and accurate data are important factors in the data warehouse. In distributed database, large amount of database required quick response to the user which can enhance the query processing. A data warehouse is the collection of materialized views which are used to process a given set of queries. Materialized views are found useful for fast query processing. It can be the result of a query, subset of the rows or column or can be the join result. Instead of accessing the original database , some immediate result can store in the form of materialized view which can give the fast access to the data .There are two issues should be consider while developing materialized views i.e. space constraint & maintenance cost constraint .Because materialized view required large amount of space on to the disk, materialization of all queries or views are not possible. Therefore Materialized views selection is one of the important decisions in designing a data warehouse. This work will implement an algorithm which can select the queries based on frequently accessing by the user & materialized that query so that user can get fast access to there useful data & automatically query processing will be faster and maintenance of data in order to get up-to-date information..

Keywords: Data warehouse, materialized views, maintenance cost, query processing cost.

Introduction:

Data warehouse contains huge amount of data from various distributed, heterogeneous and operational data sources. Initially traditional data sources never used to perform complex query therefore to overcome the weakness of traditional databases, data warehouse have been developed. Data warehouse is the collection of many materialized view which are used to answer some aggregate queries. Materialized view can give fast query response rather than accessing the data from base table. Therefore problem generally occur which views or queries should materialized in order to get low processing cost & low maintenance cost.

Materialization of all views/queries consumes larger space on to the disk therefore we need to select some of the views/queries which are beneficial to us. This work focus on two main things while selecting the views/queries for materialization i. e. frequently accessing queries and space occupied by the data that is fetched after applying a query. After considering this two main constraint an

Algorithm will select those views/queries which contain maximum benefit per unit space. It means it will select those views/queries which can give the fast access to their useful data of the user and by which query processing will be faster.

Another part of this work will maintain the data consistently. It means as the base table will change, respective materialized view will also be changed so that user can get accurate data.

Related Work:

Harinarayan [1] presented a greedy algorithm for the selection of materialized views in which data cubes are used for query evaluation cost but in this work, storage cost and maintenance cost are not given. T.Nalini, Dr. A.Kumaravel, Dr.K.Rangarajan [2] have been proposed I-mine mining techniques by which frequently accessing query can consider after that processing cost can be calculated & space occupied by that data can also be calculated .After considering all these three parameter one threshold value will select the views for materialization. Mahip Bartere and Dr. Prashant Deshmukh [3] proposed sample views which are the index materialized view used for fast accessing data form the databases. But it is useful only those application which contain random sample form the database. Himanshu Gupta and Inderpal Singh Mumick [4] developed an algorithms to select a set of views to materialize in a data warehouse in order to minimize the total query response time under the constraint of a given total view maintenance time. Chuan Zhang and Jian Yang [5] proposed a completely different approach, Genetic Algorithm, to choose materialized views and demonstrate that it is practical and effective compared with heuristic approaches. B.Ashadevi,Dr.R.Balasubramanian [6] implemented an algorithm in which query which has highest weight than the other are used for materialization and it can delete an existing materialized view which are no longer useful as well as maintenance is also well defined in this work. Sanket Patel and Deepak Dembla [7] have proposed the algorithm which select the views for materialization and it select the best node to distributed materialized view. Imene Mami and Zohra Bellahsene [8] have proposed work in which, it select views based on workload and processing

cost and it implemented an algorithm which calculated updating of a materialized view, from which those views which are updated frequently will be deleted.

Proposed Work:

This work focus on two main things while selecting views for materialization that are frequently accessing queries and total disk space. The problem of utilizing the limited resources disk space or maintenance time to minimize the total query processing cost comes under the materialized view selection with resource constraints.
Following algorithm will implement the above

Priority Based Algorithm:

Materialized view consumes large amounts of disk-space due to their large sizes that usually reach hundreds of terabytes. Therefore it is not possible to materialize all the views. This work will implement an algorithm which materialized views with highest frequency and continue it till the total disk space will be fulfilled.

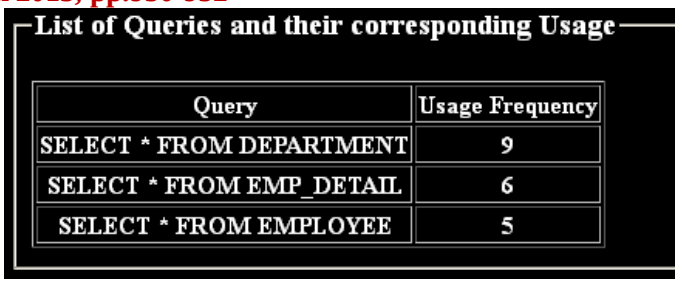
Algorithm:

1. Initialize disk size d.
2. Cerate array y for frequently accessing views.
3. Create a target array z for materialized views.
4. Repeat
5. Retrive x=[Current_view,Goal_view] from y.
6. if i==queue_size.
7. then return(Goal_view)
8. Add Current_View into z.
9. If union(Current_viewsize,Next_viewsize)<= d
10. Add Next_View into z.
11. Until queue is empty.

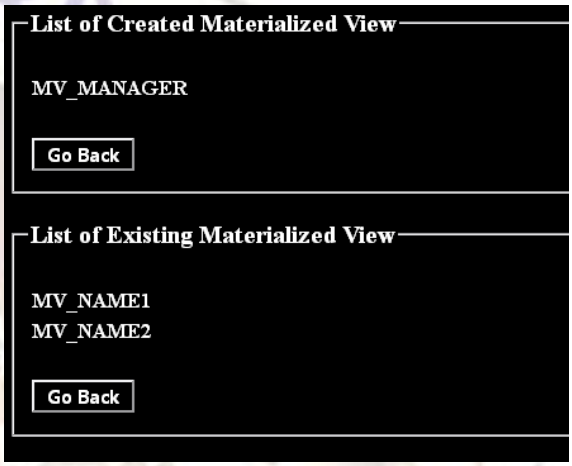
In the above algorithm the work will consider both constraint: the total disk space and frequently accessing queries by which it can improve query performance . Therefore it will be useful in huge amount of database.

Results:

Initially this work counts the frequency of each query and creates the materialized views of those queries which frequency goes beyond the threshold value.
This gives the following result of an algorithm:



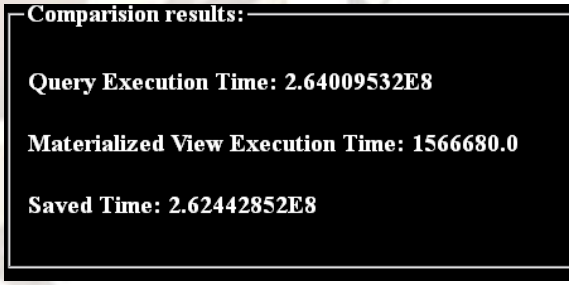
After this it created the materialized views of those query which frequencies are more than the threshold value.
Following are the materialized views:



This materialization based on highest frequencies of an queries.

After creation of materialized view whenever user will fire the same query, data will be fetched directly from the materialized view which will take less processing time.

This can be shown as follows:



Conclusion:

The selection of views to materialize is one of the most important issues in designing a data warehouse. So as to achieve the best combination of good query response where query processing cost should be minimized in a given storage space constraints. The space constraint is the most important factor while selecting the views to be materialized. The Proposed approach will implement such an algorithm that will select the

materialized views to minimize the space & storage constraint. The materialized views are selected to give minimum query response time.

References:

[1] 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.

[2] T.Nalini, Dr. A.Kumaravel, Dr.K.Rangarajan , "An Efficient I-MINE Algorithm for Materialized Views in a Data Warehouse Environment" , International Conference on Computer Science, Vol.8, Issue 5, No 1, September 2011.

[3] Mahip Bartere and Dr. Prashant Deshmukh, "Optimization of Query Processing Time Base on Materialized Sample View ", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 1222-1228

[4] H. Gupta, I.S. Mumick, Selection of views to materialize under a maintenance cost constraint. In Proc. 7th International Conference on Database Theory (ICDT'99), Jerusalem, Israel, pp. 453–470,1999.

[5] C. Zhang and J. Yang, "Genetic algorithm for materialized view selection in data warehouse environments," Proceedings of the International Conference on Data Warehousing and Knowledge Discovery , LNCS, vol. 1676, pp. 116–125, 999.

[6] B.Ashadevi and Dr.R.Balasubramanian, "Optimized Cost Effective Approach for Selection of Materialized Views in Data Warehousing", JCS&T Vol. 9 No. 1,April 2009.

[7] Sanket Patel and Deepak Dembla, "An Approach for Selection and Maintenance of Materialized View in Data Warehousing", International Journal of Information Sciences and Application.ISSN 0974-2255 Volume 4, Number 1 (2012).

[8] Imene Mami and Zohra Bellahsene, "A Survey of View Selection Methods", SIGMOD Record, March 2012 (Vol. 41, No. 1).

[9] MR. P. P. Karde, DR. V. M. Thakare, "Materialized View Selection Approach Using Tree Based Methodology",ISSN:0975-6698-NOV 09 to Oct 10-vol 1.