Acceleration Of Query Response In Distributed System Using Materialized View Approach

Pooja G.Dhawane
Department CSE ,R.T.M.N.U.
poojadhawane@gmail.com

Prof. Garima Singh makhija
Department CSE ,R.T.M.N.U.
garima11makhija21@gmail.com

Abstract—*A data warehousing technology is nothing but essential key pieces of information which are used by industries for the business strategy formulation and implementation. The success of data warehouse depends on appropriate data delivery method. The delivery process should be quick and accurate. This can be achieve by enhancing the query execution i.e, by minimizing the query response time. Performance of query can be improved by different approaches like hybrid integration system where most recent data is stored at mediator i.e. common access point to distributed system ,optimization of query, use of proper data structure, reduce complexity of query etc. leaving all alternatives, we are planning to use materialized view approach. Materialized view is used to store the result of queries. We are going to store the result of queries based on the frequency of queries fired by the user. The queries having high frequencies i.e, based on the user demandable queries materialized view is designed. This will save the time of revisiting the base table whenever any query is fired frequently and also saves the time of cost of operations like join, aggregations etc.*

Keywords— Materialization, query frequency, optimization, data cluster, data warehouse.

1. INTRODUCTION

Data warehouse is an approach to the integration of data from multiple, possibly very large, distributed, heterogeneous databases and other information sources. [1]

Data warehousing is the collection of huge amount of data for querying and analysis. Fundamental requirement for the success of a data warehouse is ability to make decision with both accurate and timely consolidated data also fast query response time.

In existing system [2] the attributes which are most probably to extract where placed at the common access point called mediator to the distributed system. This has reduced the access time for the getting the data from original source of data.

But if the attributes placed at the mediator are not proper then performance of the query will not be up to the mark.

The primary intent of this research is to develop a system which will reduce the execution time for query i.e, getting the response from database as early as possible which in turn improves the performance of queries. We are going to take the help of materialized view. To avoid accessing the original data sources and increase the efficiency of the queries fired on data warehouse, some results in the query processing are stored in data warehouse.

2. RELATED WORK

Effective query processing using materialized view approach has been studied and implemented by many researchers by their own ways to achieve query optimization so as to decrease processing time.

In Hybrid integration system [2] some part of data were placed on mediator i.e. common access point to data warehouse. The attributes of interest are extracted from various data sources then possible values are assign to them and this information is placed on mediator. But if the attributes chosen are in appropriate then query performance will not be at the mark .Thus, we are trying to eradicate this problem using materialized view which directly stores the result of queries executing frequently

Hema Botre and M. S. Choudhari [3] presented a Priority based algorithm for materialized view selection.this algorithm selects the queries based on the priority associated with queries for the materialization.

Harinarayan proposed greedy algorithm [4] using data cubes to select the materialized view but this work does not consider the storage space required.

Sanket Patel used Tree based algorithm [5] for the selection & maintenance of materialized view. The decision of which data set should be materialized 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.

Y.D.Choudhary [6] proposed a novel CBFSMV algorithm for selection of materialized view using query clustering strategy that reduces the time as compared to response time for actual database.

In [7], authors have implemented dynamic adjustment for static materialized view selection algorithm that is EMVSDIA (Efficient Materialized view selection Dynamic Improvement algorithm).Experiments demonstrates EMVSDIA has excellent implementing efficiency and reach to expected efforts.

B. Ashadevi, Dr.R. Balasubramanian [8] implemented an algorithm that is projected for choosing the views to materialize on basis of their weight acquired in the query set. A greedy algorithm is used to incorporate the maintenance cost and storage constraint in the selection of data warehouse Materialized View. It reduces comp-lexity of algorithm compare to previous algorithm [9].
Hybrid algorithm is combination of Evolutionary algorithm and Heuristic algorithm. The result of algorithm show that applying an evolutionary algorithm to either global processing plan optimization or materialized view selection for a given global processing plan can reduce total query and maintenance cost. Excessive computation time of pure Evolutionary algorithm and unsatisfactory quality of solution of Heuristic algorithm can be overcome by using Hybrid algorithm[10].

3. IMPLEMENTATION FRAMEWORK

After studying different research paper and analysing different algorithm to select MV. By keeping in mind the goal of our work to enhance performance of query in data warehouse using MV. We have divided the work into four parts, as shown below.

3.1 Creation of data warehouse
3.2 Data search by attribute
3.3 Creation of Materialized view
3.4 Search using MV

3.1 Creation of data warehouse

In this phase we collect raw data i.e,dat file
is used. This raw data is loaded in temp.table. After cleaning the data ,data warehouse is created. Now data is available by executing query on them. once the data warehouse is created it is used to create cluster data warehouse .These clusters are distributed on different systems. Clusters are formed on attributes.

3.2 *Data search by attribute*

We fire the queries on data warehouse and record the time needed to search the record based on attribute.

3.3 *Creation Of Materialized View*

In this phase MV is created by firing the queries .Each of search query is added in .txt file as per that MV is created. The size of MV depends on what type of database is used. It works like cache memory.

3.4 *Search using MV*

User fires the query. If data is available in MV user does not visit the base table if miss occurs then record is search in original data source. The time required to search the data using MV get reduced as visiting to base table is avoided.

4. EXPERIMENTAL RESULTS

This section shows the running experimental results that are carried out using materialized view schema by applying steps A, B, C, D. The sample queries are fired on database and following outcomes are obtained.

.

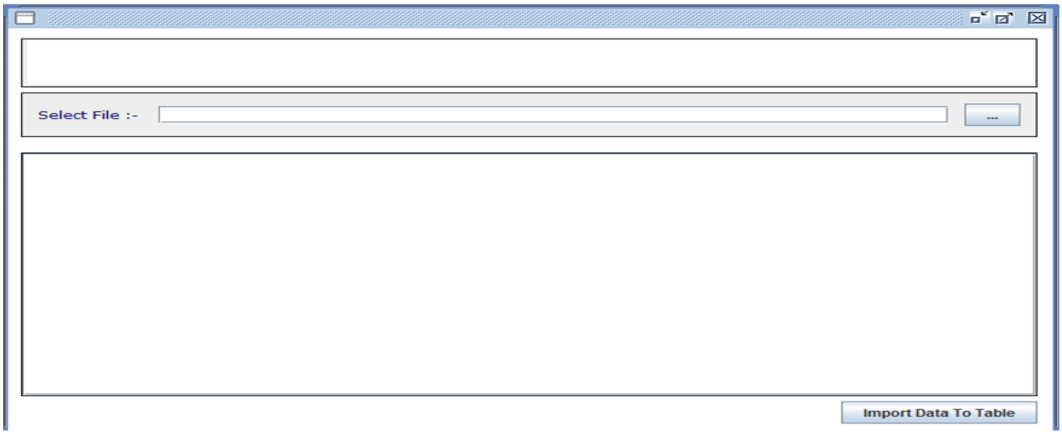
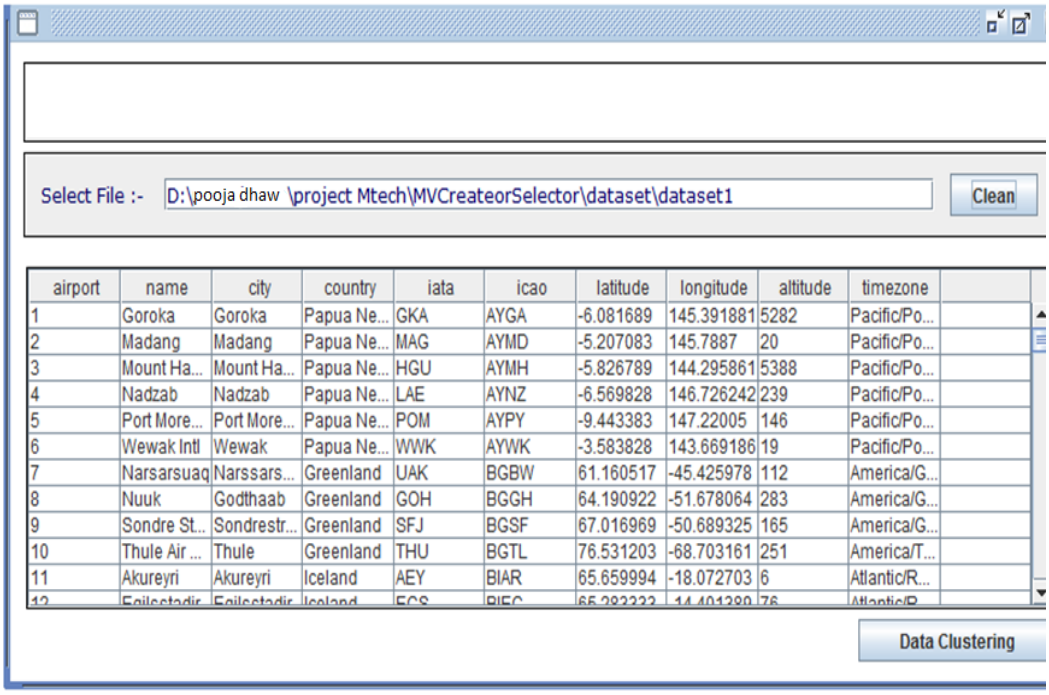


Fig.1.selection of Raw Data



airport	name	city	country	iata	icao	latitude	longitude	altitude	timezone
1	Goroka	Goroka	Papua Ne...	GKA	AYGA	-6.081689	145.391881	5282	Pacific/Po...
2	Madang	Madang	Papua Ne...	MAG	AYMD	-5.207083	145.7887	20	Pacific/Po...
3	Mount Ha...	Mount Ha...	Papua Ne...	HGU	AYMH	-5.826789	144.295861	5388	Pacific/Po...
4	Nadzab	Nadzab	Papua Ne...	LAE	AYNZ	-6.569828	146.726242	239	Pacific/Po...
5	Port More...	Port More...	Papua Ne...	POM	AYPY	-9.443383	147.22005	146	Pacific/Po...
6	Wewak Intl	Wewak	Papua Ne...	WWK	AYWK	-3.583828	143.669186	19	Pacific/Po...
7	Narsarsuaq	Narssars...	Greenland	UAK	BGBW	61.160517	-45.425978	112	America/G...
8	Nuuk	Godthaab	Greenland	GOH	BGGH	64.190922	-51.678064	283	America/G...
9	Sondre St...	Sondrestr...	Greenland	SFJ	BGSF	67.016969	-50.689325	165	America/G...
10	Thule Air ...	Thule	Greenland	THU	BGTL	76.531203	-68.703161	251	America/T...
11	Akureyri	Akureyri	Iceland	AEY	BIAR	65.659994	-18.072703	6	Atlantic/R...

Fig.2. Cleaning of Raw Data

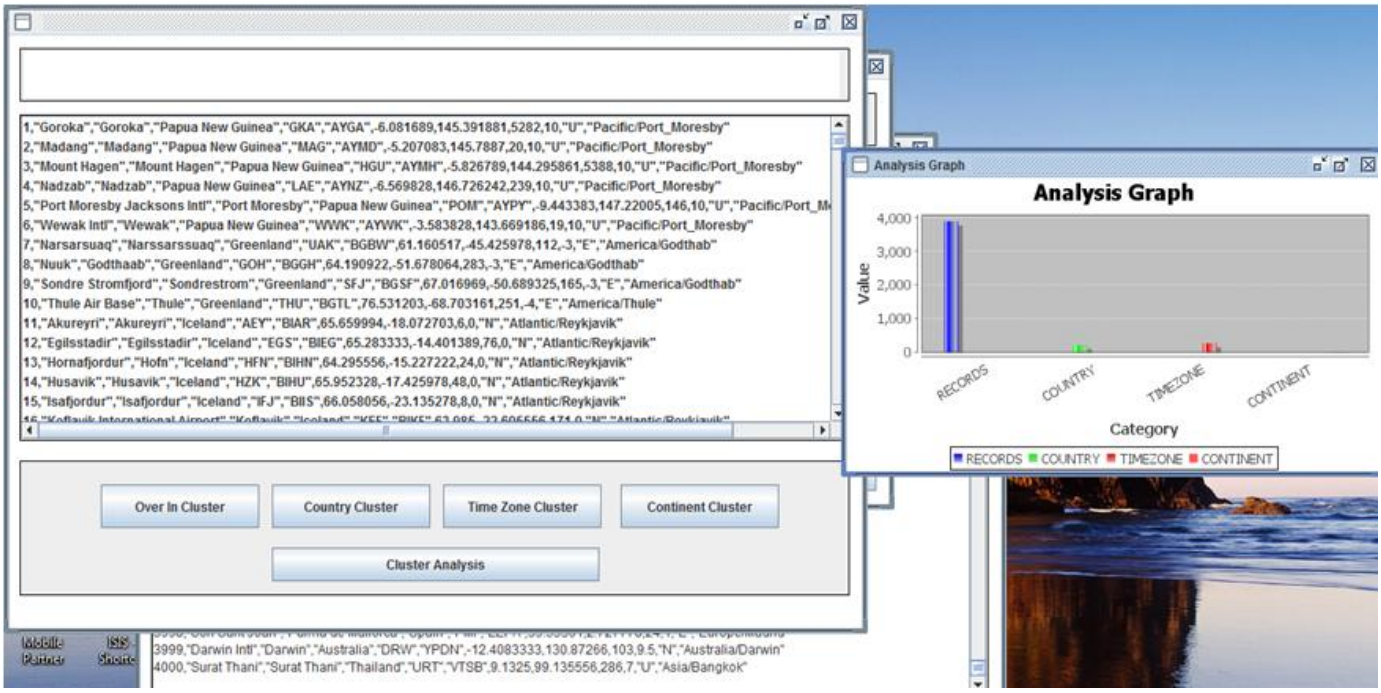


Fig.3.Data Clustering on Attribute Country, Time zone, Continent

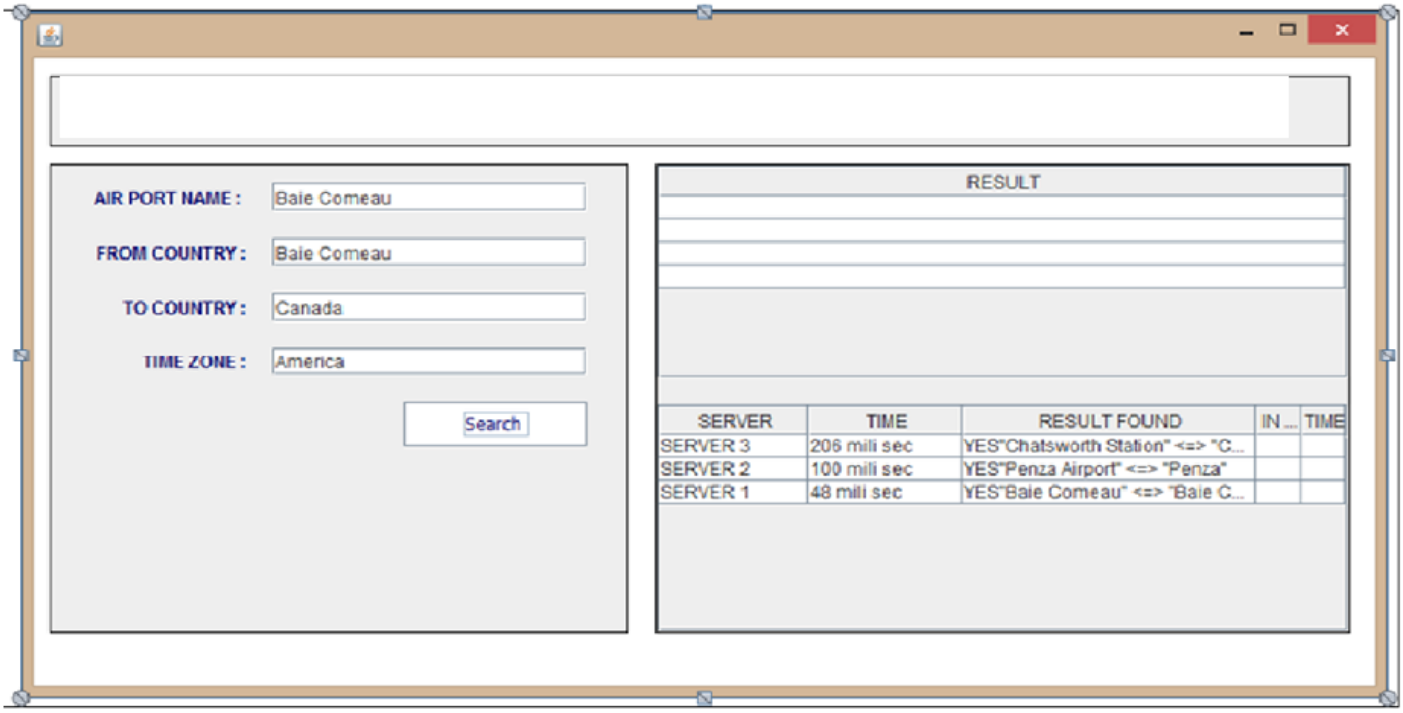


Fig.4.Query Posted On Data warehouse without MV and Time is Recorded

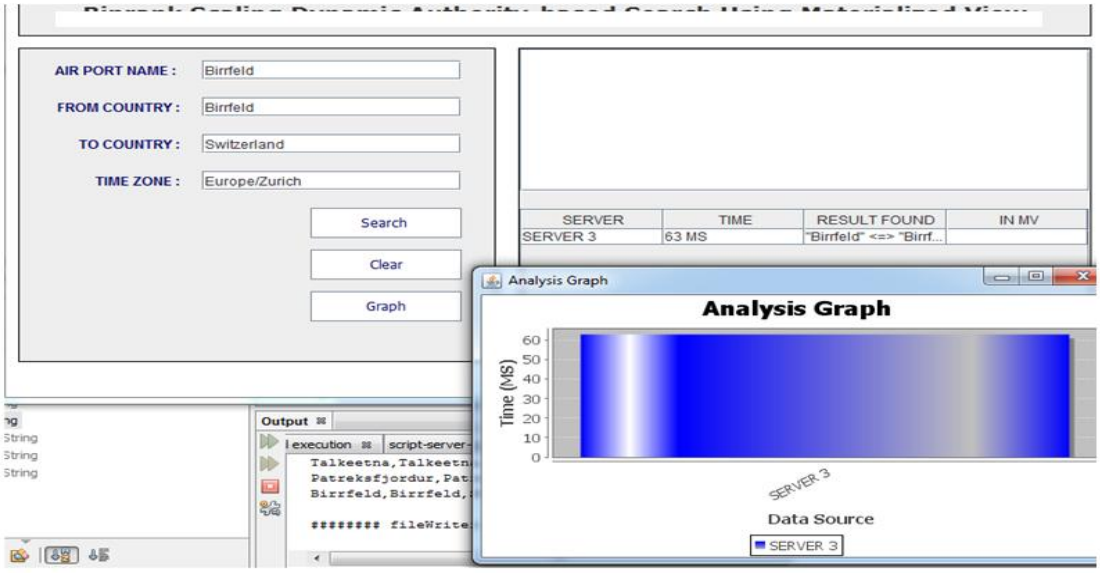


Fig.5. Graphical Representation of Search Time To Retrieve Data From Base Source

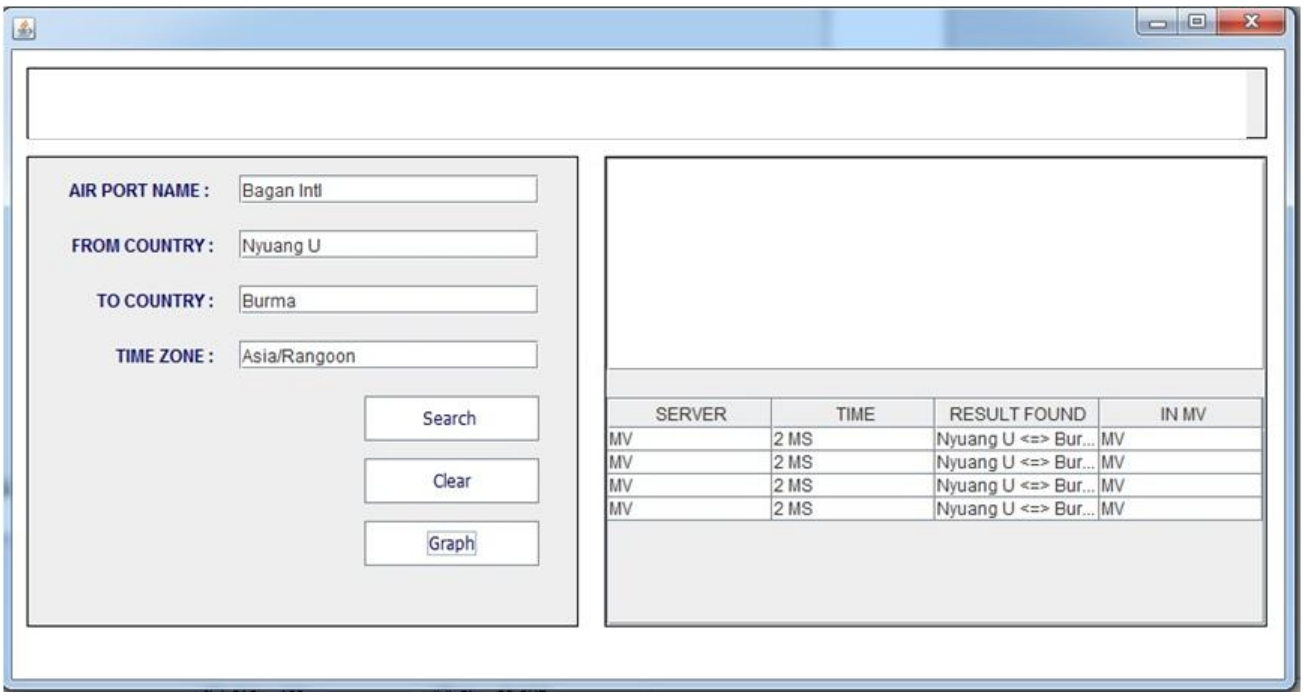


Fig.6.Query is Posted Using Implementation of MV and Time is Recorded

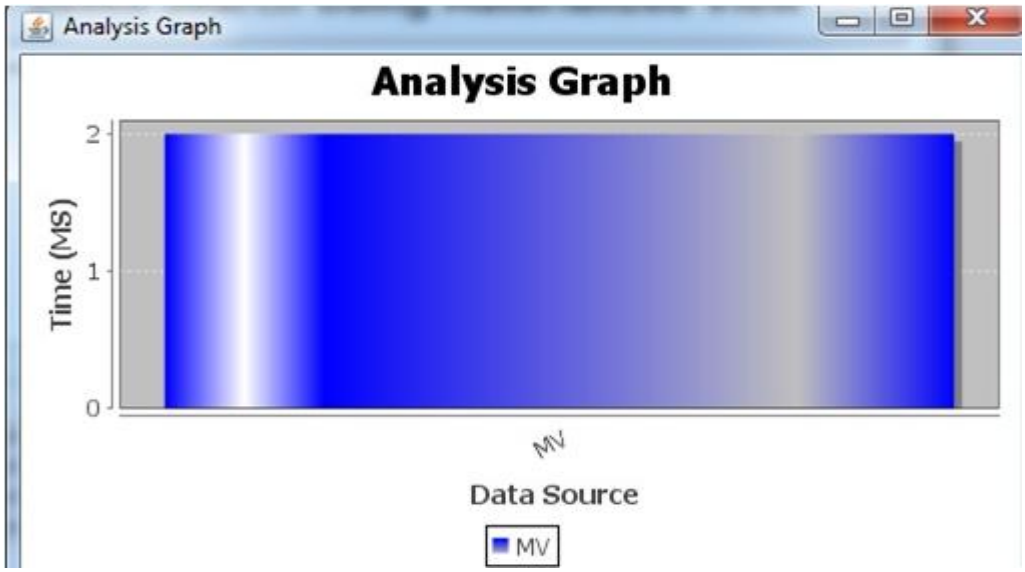


Fig.7.Graphical Representation of Search Time To Retrieve Data Using MV

5.CONCLUSION

From the above experimental result , we can see that with the help of materialized view the time to visit the base table is saved. Hence, the performance of query is improved to 15 to 16 times. Because MV is having the result of frequently fired queries.

Thus, by implementing the materialized views in distributed system, we are trying to get the combination of good query response time from which query processing cost should be minimized.

6.REFERENCES

[1] *W.H. Inmon, Building the data warehouse, 2nd ed, John Wiley and Sons, Canada, ISBN:0471 -14161-5, 1996.*

[2] Samir Anter, Ahmed Zellou, Ali Idri "Hybrid Integration System", 2013, 5 th IEEE International Conference on zComputer Science and Information Technology (CSIT).

[3] Hema Botre, M.S. Choudhari, "Design an Algorithm For Materialized view Selection",2013, IJERA

[4] V. Harinarayan, A Rajaraman, and J. Ullman. "Implementing data cubes efficiently". Proceedings of ACM SIGMOD 1996 International Conference on Management of Data, Montreal, Canada 1996.

[5] .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).

[6] Yogeshree D. Choudhari, Dr. S. K. Shrivastava, "Cluster Based Approach for Selection of Materialized Views, Volume 2, Issue 7, July 2012.

[7] Zhou Lijuan , Ge Xuebin and Wang Linshuang,Shi Qian, 'Efficient Materialized View Selection Dynamic Improvement Algorithm' , 6th International Conference on Fuzzy Systems and Knowledge Discovery, Vol.7, pp.294-297, 2009.

[8]] 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.

[9] Gupta, H. & Mumick, I. Singh, "Selection of Views to Materialize in a Data Warehouse" IEEE Transactions on Knowledge and Data Engineering, 17(1), 24-43, 2005.

[10] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, no. 2, article 9, 2009.