CHAPTER - 5
CONCLUSIONS / IMP. FINDINGS

In today's scenario data warehouse plays a crucial role in order to perform important operations. Different indexing techniques has been used and analyzed using different types of queries on different size of datasets in data warehouse in order to perform operation in efficient manner. Our grouping model provides better performance than cluster and non cluster indexes. This model analyzes the performance of query on data warehouse. From this work it is found that our model works as twice as fast as that of other previous defined models. It is also found that the time needed for query execution of previous model is high when compared to our model. It is also found that the running time get decreased on an average of 4 times when the working on same data.

Data warehousing is the and most trustworthy technology used today by corporatist for scheduling, forecasting, and management. After the evolution of the concept of data warehousing during last two decade it was thought that this technology grow at a very rapid pace but unfortunately it's not the reality. Many research has been done in this ground regarding design and development of data warehouses and a lot still requirements to be done but one area which requires special awareness from research community is data warehouse maintenance.

A major reason for data warehouse project failures is pitiable maintenance. Without suitable maintenance preferred results are nearly

unfeasible to attain from a data warehouse. Unlike operational systems data warehouses need a lot more maintenance and a support team of qualified professionals is needed to take care of the issues that arise after its implementation including data extraction, data loading, management of data flow on network , coordination and communication, training, education and documentation and some other related tasks. To take out all these functions and processes a competent team of full time trained experts is required who can competently and continuously take care of all the data warehouse maintenance problems in a well-timed manner.

During my empirical study I have validated seven major areas which have to be managed in a timely manner for a better and improved performance. All these concepts are closely related to the theoretical findings and the empirical findings with very little differences. These are:

5.1 Query Performance
5.2 Coordination and Communication
5.3 Education, Training and documentation
5.4 Help and support
5.5 Management of data flow on network
5.6 Loading and Cleansing Data
5.7 Software and Hardware
5.8 Materialized view

5.1 Query Performance

In today's scenario data warehouse plays a crucial role in order to perform important operations. Different indexing techniques has been used and analyzed using different types of queries on different size of datasets in data warehouse in order to perform operation in efficient manner. Our grouping model provides better performance than cluster and non cluster indexes. This model

analyzes the performance of query on data warehouse. From this work it is found that our model works as twice as fast as that of other previous defined models. It is also found that the time needed for query execution of previous model is high when compared to our model. It is also found that the running time get decreased on an average of 4 times when the working on same data.

5.2 Coordination and Communication

The importance of communication across different business functions and departments is well known in the IT implementation literature. According to one author on IT project management, **'communication is the oil that keeps everything working properly'.** The communication process also continues along with the training program. The coordination and communication between process keeps the business users and IT users in contact with each other to have exchange of views, suggestions and any guidance towards better performance of a data warehouse. The case study shows that most business users are reluctant to adopt technology to carry out their work, therefore pursuing a business user to use data warehouse is inevitable. To pursue business users in using data warehouse the communication and training program is a must.

Data warehouse help and support team members can continue the coordinate and communication between program by keeping themselves in close contact with the business users and exchanging views on the present and proposed performance of data warehouse etc. Some familiar parties or gathering could be arranged the communication and parallelism coordinate the gap between business community and the IT community. Information about the data warehouse in the form of documents, emails or presentations should be regularly sent to the data warehouse users.

If top management is not strongly communicate with Mid Level

Management, or does not actively participate in the implementation, the implementation has a high likelihood of failure. The implementation of any project must be viewed by top management as a transformation in the way the company does business. They should assure the employees about their jobs, clear any doubts and explain why this system is a necessity for the organization.

The CEO or some senior level manager should sponsor any project in order to demonstrate the management commitment. Apart from top-level commitment, Mid-level management commitment is also required for the implementation. Middle levels of management must participate when determining the detailed implementation plans. This was ensured that everyone's interests and concerns are considered before final decisions are made. Mid-level managers should have hands-on responsibility and authority for the detailed aspects of the implementation. The lack of top/middle management communicate inevitably lead to failure. If the implementation is treated as simply an IT project, then system never realize its full capabilities. In such cases, it is similar that the technology is deployed in a vacuum, business processes not be properly reengineered and aligned with the software requirements, and staff resist using it and indirectly it would lead to failure.

5.3 Education Training and Documentation

My empirical study prove that the initial and the major significant part of a data warehouse maintenance plan is the training of users.

Supplementary I accomplished from the empirical study that the education plan should be implemented according to the capabilities and skills of the users. Different levels of training should be carried out for different groups of users.

The business users can be trained in-house provided qualified professionals are present in the data warehouse support team which is a must while advanced technical training for support team can be carried out by the vendor or some outside party. The advanced training includes lectures on logical data modeling and Loading and Cleansing Data functions.

The training program gives the users of data warehouse an insight into the qualities and capabilities of a data warehouse and teaches them the methods to benefit from it. Often the data warehouse projects fail because the users don't know how to use it according to the business needs. No one is going to use the data warehouse until they know how to use it, especially the business users who are more comfortable in receiving reports in a paper form instead of using computers for this purpose.

The very important consideration in the planning of user access is education, training and documentation. Information System management should pay serious attention to training users on data access and analysis tools. Users should also be trained in the layout and structure of the current data ware house, as well as in the use of metadata resources.

Users should be encouraged to learn SQL, spreadsheets, and other statistical data analysis packages to help them directly meet their specific data needs through the data warehouse. Such education and training programs should be regularly and repeatedly conducted at time slots convenient to users.

Organizations that provide visibility and rewards, such as letters of recognition and certificates, cash awards, salary raises, and mention in the in-house newsletters, would find it easier to motivate users to learn the new tools.

A high degree of inter- action with users in training programs can lay the foundation for obtaining regular user feedback and suggestions for future refinements. The usefulness and quality of a DW and its contribution in enhancing business performance ultimately depend on the developers' abilities to make the DW environment user friendly. This must be viewed as an ongoing process requiring continued financial commitment rather than a one-time effort.

5.4 Help and Support

My empirical study showed that services of help and support play an important role in taking valuable output from the data warehouse. Help and Support is always required in any information system, same is the case with a data warehouse but here the support is needed 24 hours a day. Some of the processes which require presence of support staff to rectify any problem. There are mechanisms implemented where the support staff is notified of any problem by the help of a SMS to his/her mobile phone. The support staff after receiving the message takes necessary steps to rectify the problem.

Support staff can also be contacted by data warehouse users if they had some problem or difficulty in using the system. There should be some sort of formal or informal help desk created solely for this purpose depending on the size of organization and the data warehouse. The help and support counter personnel should be available at all times through any means such as by phone, by email or in person. The support team logs all the problems found and reported by users for later references and user community can also benefit from previous experiences.

Users can contact support team in any problem whether related to security and sign in, data access to detailed or aggregated data, query processing, report

generation, front end tools usage etc. The support team also points out if there is any loop hole or problem area within the data warehouse that should be addressed. Apart from help desk each data warehouse support team develops its own problem management process in which they understand local language of the employee and solve their problem in efficient manner.

The process defines necessary routines and instructions to counter any problem found in the warehouse. If the problems found in the data warehouse are not addressed at the right time, this leads to performance shortfalls, and usability and availability issues in near future. Thus help desk and problem management play a key role in improving data warehouse performance and getting the desired output from it.

5.5 Management of data flow on network

Management of data flow on network also plays its part in improving data warehouse performance. From the empirical study I concluded that by having a fast and reliable network user queries get a much shorter response time especially in a distributed data warehouse. There should be some specialized persons responsible for managing the network in the organization. As more and more users start using the data warehouse the load on the network also increases and response time becomes longer and longer. The network monitoring personnel's can use some specialized tools for monitoring network performance. The features these tools should provide are listed in the theoretical part.

Another approach for network management could be using the services of some specialized company in this field. Such firms can apply network planning, design, implementation, management and monitoring services either remotely or on site.

These objects are the hardware, software, and virtual components of the network. They form a composition hierarchy whose underlying structure is fairly standardized.

When the network encompasses multiple types of hardware, uses complex switching algorithms such as Asynchronous Transfer Mode (ATM), or uses sophisticated industry standards, this hierarchy may be asymmetric and of indeterminate depth. Storing this configuration hierarchy in a database requires propagating operations from higher levels to lower levels.

For instance, removing a switch from the network must also remove all of that switch's boards, all the boards' ports, etc. If an operator wanted to modify the configuration of a switch, the database needs to lock all of its subcomponents to prevent a different operator from modifying the parameters of a board within the switch. Also, nearly every operation performed by network management applications requires configuration data. Therefore, the database storing this information must provide high concurrency, fast navigation, and high availability.

5.6 Loading and Cleansing Data

The empirical study showed that loading and cleansing functions needed to be carried out by a competent and trained professional team. This team is leading by an expert. It’s the responsibility of architect of team to devise a comprehensive and effective loading and cleansing of process to load the data warehouse. The expert ensures that the loading and cleansing of processes have strength and survival. The architect works in close coordination and communication with the business users and identifies which data and at

what level of detail is required. The architect has a clear understanding of the company's business and knows what type of data is exactly required. The data warehouse not addresses reporting requirements, until and unless it has the data that is useful. If the data is of no use for the business, there is no need storing it in the warehouse.

Thus architect has this very sensitive responsibility of selecting the right and useful data for loading into the data warehouse. After identifying data that needs to be extracted from the operational systems trained professional team transforms it into a format acceptable for the data warehouse and cleanses it as well by eliminating useless data. The data is finally loaded into the warehouse. During all these tasks it is the responsibility of trained professional team to make sure that all the operational systems and the data warehouse are available to users.

Accurate data is an absolute requirement for any system to function properly. If inaccurate data is entered into the common database, the erroneous data may have a negative impact throughout the enterprise. Inaccurate data can lead to errors in market planning, production planning, material management, and financial planning. If a company with inaccurate data just forges ahead under the assumption that data errors be corrected when they are spotted, then the project lose credibility. This encourages people to ignore the new system and continue to run the company under the old system.

5.7 Software and Hardware

Upcoming planning for hardware and software possessions for the data warehouse is essential for taking maximum output from it. In the beginning as the data warehouse usage is less, fewer resources are required, but as time

passes and data warehouse starts becoming popular within the users, more hardware and software resources are required for a smooth running. It is the responsibility of the data warehouse support team to keep an eye on the data warehouse trends within the organization and develop a strategy for calculating the required hardware and software resources for the future.

Possible hardware solutions include installing bigger and faster processing hardware and storage facilities, parallel processors, specialized multidimensional database servers, and distributed data warehouses. However, a hardware approach to this problem represents a very costly alternative requiring new technologies and skills and, therefore, should be the last option.
One possible software-based solution is the use of tools such as OLAP (online analytical processing), where the queries are tuned for multidimensional analysis along preplanned dimensions. Another approach is partitioning, where the data is segmented into logical areas. Such partitioning can reduce search efforts. An unpopular and unlikely option is to regulate the kinds of queries that users can put to the DW through query trapping. Unfortunately, these software solutions may not solve performance and access problems in ad hoc querying. An attractive alternative is to build inverted indexes in addition to the traditional B-tree indexes found in relational databases. Inverted indexes contain pointers to the database as data and can be effective for even unstructured and qualitative data. Overall, a well-designed set of indexes provides an efficient solution to the problems of data access in the DW environment without requiring special hardware or major restructuring of the DW model.

Having up to date versions of software and hardware is another issue that the data warehouse support team needs to handle. For this purpose the companies can sign agreements with vendors or they could implement it in house if

they have the required competence and resources within the organization. Software updates are usually provided by the vendors. The data warehouse help and support team should be in contact with the vendors and needed to provide feedback on data warehouse performance to them. Hardware updates are inevitable when the existing hardware can't withstand the computing load and the response time increases. In such cases one needs to install/update new hardware. The updating could be in the form of ram upgrading or disk upgrading, alternatively new and latest hardware could be installed.

5.8 Materialized View

View materialization is a strategy used to provide fast answers to user queries. But it is important to have updated views whenever the base tables upon which views are built are updated. It is the responsibility of data warehouse support team to devise a flexible and optimal strategy for maintenance of materialized views. Although it is more of a design decision to provide capabilities for view maintenance but it is the user's decision to decide which views to materialize and when to refresh them.

The goal of database performance tuning is to minimize the response time of queries and to make the best use of the server's resources by minimizing network traffic, disk I/O, and CPU time. This goal can only be achieved by understanding the logical and physical structure of data, understanding the applications used on server, and understanding how the many conflicting uses of database may impact database performance.

The best way to avoid performance problems is to ensure that performance issues are part of ongoing development activities. Many of the most significant performance improvements are realized through careful design

at the beginning of the database development cycle. To most effectively optimize performance, it must identify the areas that yield the largest performance increases over the widest variety of situations and focus their analysis on those areas.

Sometimes a query can be quite difficult. Often, it is possible to get the desired result through the use of sub queries. For those cases, the in-between results are not stored in the database, but are directly used within the query. This can escort to performance issues, especially when the intermediate results have a large number of records.

The way to increase query performance in those cases is to store the intermediate results in a temporary table, and break up the initial SQL statement into several SQL statements. In many cases, you can even build an index on the temporary table to speed up the query performance even more. Granted, this adds a little complexity in query management (i.e., the need to manage temporary tables), but the speedup in query performance is often worth the trouble.

Network management applications often track a lot of information over time to analyze performance and perform billing. In order to tune the network or plan for future needs, system administrators may need to execute queries that span petabytes of this data. The database must execute these queries in a reasonable time. Also, since critical automated systems update this data, the database must allow read/write concurrency. Otherwise network service would effectively freeze whenever an administrator made such a query.