Performance Tuning Mechanisms for Implementation a Successful Data Warehouse

Vishal Gour¹, Prof. S.S.Sarangdevot²
¹Research Scholar, Dept. of Computer Science, Suresh Gyan Vihar University, Jaipur, India
²Director, Dept. of Computer Science and Information Technology, J.R.N. Rajasthan Vidyapeeth University, Udaipur, India
{goar.vishal@gmail.com, dr.sarangdevot@gmail.com}

Abstract - Data warehouse is the design and implementation of process, tools to manage and deliver entire, timely accurate and logical information for decision making. It is accepted that information is a very powerful feature that can provide significant benefits to any organization. Data ware house are computer based data information system that optimized database query and reporting tool because of their ability to analyze data often from disparate database and in interesting ways. They are a way for managers and decision makers to extract information quickly and easily in order to answer question about their business. Data ware house is not magic – they require a great deal of very hard work. In many cases data ware house assignment are viewed as a stopgap evaluate to get users off to backs or to provide something for nothing. But data ware house require alert management. A data ware house is good investment only if end users actually can get important information quicker and cheaper they can use modern technology.

I will give concept about how to perform and how to get the word out to the end user community and management has to recognize that the maintenance of any other mission critical application. So management has to take decision for very high level of maintenance. Because previous researchers have focused on methodological, data related, operational, educational and technical issue of data warehouse implementation. There is require explore studies in which discuss organizational, project-related and environmental dimensions regarding implementation of data warehouse technology in general and in companies then we came to know that current scenario require the data warehouse to change.

Keywords: Data warehouse, decision making, management, maintenance, end user, critical application.

1. Introduction

Data warehouse are actually build new system environment. This new environment is kept separate from the system environment helping in the operations. The primary goal of data warehouse is use the information so that decision makers and business analyst can make queries, analysis and planning regardless of the data changes in operational database.[10] The data ware house essentially holds the business intelligence for the enterprise to enable strategic decision making. A data ware house is designed for analysis and query rather than for a transaction processing. It separate analysis from transaction and permit an organization to combine data from several sources.[7] A vibrant strategy to improve the data warehouse performance. A data warehouse is the copy of transaction data especially structured for querying, reporting and analysis purpose. The data warehouse contains copy of transaction which cannot be updated or altered by the transaction system. Data warehouse is the source of stable and integrated data designed to support decision makers and business analysts. [1] Data are acquired from various operational data stores across the enterprise. When a database management system (DBMS) parses a query it decides the best strategy to execute it based on statistics it retains about DB structure, indexes, and number of distinct values.

2. Literature Review

The art of performance tuning has always been about matching workloads for execution with resources for that execution. Therefore, the beginning of a performance tuning strategy for a data warehouse must include the characterization of the data warehouse workloads. To perform that characterization, there must be some common metrics to differentiate one workload from another.[8] If the data warehouse has that cannot be transformed properly, it is important for the data warehouse transformation process to use intelligent default values for the missing or corrupt data. It is also important to devise a mechanism for users of the data warehouse to be aware of these default values [3].

Performance issues in data warehousing are centralized around access performance for running queries and incremental loading of snapshot changes from the source systems.[9] The following six concepts can be considered for a better performance:

1. A communication process also offers the data warehouse team to measure progress and identify and resolve issues before they become a serious problem. The communications program provides the business components with increased capabilities and functions. Communication is the medium or the process by which one can convey or express his thoughts, views and feelings. Whatever be the mode of communication, the effectiveness of the communication is very important for the success of an individual or a team.

2. The quality of employees and their development through training and education are major factors in
determining long-term profitability of a small business. Training needs to be focused on data warehouse concepts and terminologies, introduction to the organizational data, where is that located in the warehouse and how it is related to the reports or systems user already is using, the mechanics of using the tool. It is important for people to understand basic navigation within the tool.

3. The Call Center is an important division for any organization as it serves as the primary interaction point between customers and the company. In many situations, it is the only interaction point and therefore, responsible for the customer’s experience and satisfaction. Due to this heightened level of importance, it is critical that the contact handling process is conducted both efficiently and effectively. A Call Center acts a coordinating body for not only collecting and logging problems with the data warehouse environment but also determining where future requirements may lay.

4. Modern communication networks create large amounts of operational data. Including traffic and utilization statistics and alarm/fault data at various levels of detail. The networking hardware is proliferating with LANs, WANs, hubs, routers, switches and multiplexers. Leaving behind all this is the next stage – users wanting to access internet based data sources along with the corporate data, requiring even greater bandwidth and network management resources. Managing this environment is one big challenge, capacity planning for the future is another. If the data warehouse team is not quite good in networking technology than there should be at least one person in the organization who understands technology.

5. Extract, transform and load (ETL) is the core process of data integration and is typically associated with data warehousing. ETL tools extract data from a chosen source, transform it into new formats according to business rules, and then load it into target data structure. After it is collected from multiple sources (extraction), the data is reformatted and cleansed for operational needs (transformation). Finally, it is loaded into a target database, data warehouse or a data mart to be analyzed.

6. A view can be materialized by storing the tuples of the view in the database. Index structures can be built on the materialized view. Consequently, database accesses to the materialized view can be much faster than recompiling the view. [12] A materialized view is thus like a cache [a copy of the data that can be accessed quickly.] a materialized view provides fast access to data; the speed difference may be critical in applications where the query rate is high and the views are complex so that it is not possible to recomputed the view for every query.

The performance of a Data Warehouse is largely a function of the quantity and type of data stored within a database and the query/data loading work load placed upon the system. When designing and managing such a database there are numerous decisions that need to be made that can have a huge impact on the final performance of the system.[4] As the demand on a data warehouse changes, a lot of changes needed to be carried out to keep the performance graph in a positive direction. Some training courses needed to be introduced, some changed are needed for Help counter, some indexes become obsolete and others need to be created, some aggregates are no longer referenced and others need to be evaluated, and the limits on parallel processing must be assessed and adjusted to fit the current demand.[11] These and other tuning tasks should be carried out periodically to keep data warehouse performance smooth and constant.

3. Methodology

This research is carried out using a case study approach. The case study is conducted at Bharat Sanchar Nigam Limited (BSNL). Bharat Sanchar Nigam Ltd. formed in October, 2000, is World’s 7th largest Telecommunications Company providing comprehensive range of telecom services in India: Wire line, CDMA mobile, GSM Mobile, Internet, Broadband, Carrier service, MPLS-VPN, VSAT, VoIP services, IN Services etc. Presently it is one of the largest & leading public sector units in India. BSNL is the only service provider, making focused efforts and planned initiatives to bridge the Rural-Urban Digital Divide ICT sector. In fact there is no telecom operator in the country to beat its reach with its wide network giving services in every nook & corner of country and operates across India except Delhi & Mumbai. Whether it is inaccessible areas of Siachen glacier and North-eastern region of the country. BSNL serves its customers with its wide bouquet of telecom services.

A research model is developed to investigate the factors affecting DW success. This model serves as a guide in data gathering regarding DW success. The nature of case study research is such that employing a guide to data gathering can provide a focus without inhibiting unnecessarily the discovery of unexpected phenomena.

3.1 Research Model

Researchers have investigated the success of information systems (IS) in numerous ways [6], such as by measuring the satisfaction of users, service quality, and
the perceived usefulness of specific applications. A DW represents significant new opportunities for an organization to utilize all available internal and external information in an effective manner to help create and sustain a competitive advantage. Our recommendations for the successful planning of a DW project include starting with a careful alignment of the data warehouse objectives with the needs of the organization and defining an adequately narrow focus to implement in the first phase, as well as defining database architecture appropriate to the task. A major reason for data warehouse project failures is poor maintenance. Without proper maintenance desired results are nearly impossible 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 deployment including data extraction, data loading, network management, training and communication, query management and some other related tasks. To carry out all these functions and processes a qualified team of full time skilled professionals is required who can efficiently and constantly take care of all the data warehouse maintenance issues in a timely manner.

The communication process also continues along with the implementation factors drive implementation success that leads to system success for a DW implementation. The performances factors are collecting from the literature walk around. These are clustered into the primary drivers of Effective Communication, Training, Education and Documentation, Help, Support and Call Center, Network Management, Software and Hardware source systems, Extract, Transform and Load process and Materialized View.

The drivers, for which metrics can be established and monitored, impact the DW implementation success. Performance success is subdivided into three key components. These include success with Business, Development, and Technological issues that arise during the lifetime of the DW project. Those successes influence the data quality and structure quality, which guide the system success culminating in the supposed success of the overall DW implementation.

The communication process also continues along with the training program. The communication process keeps the business users and IT users in contact with each other to have exchange of views, suggestions and any guidance towards enhanced performance of a data warehouse.

Data warehouse support team members can continue the communication 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 informal parties or get-togethers could be arranged to bridge 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 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.

Network management also plays its part in improving data warehouse performance. From the case study we 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 person(s) 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. 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. The ETL team is headed by an ETL expert. It’s the responsibility of ETL architect to devise a comprehensive and effective ETL process to load the data warehouse. The ETL architect/expert ensures that the ETL processes have strength and endurance.

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.

4. Conclusion

A major reason for data warehouse project failures is poor maintenance. Without proper maintenance desired results are nearly impossible 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 deployment including data

**Figure 1:** Research Model for Data Warehousing Success

3.2 Research Model Description

As illustrated in Figure 1, implementation factors drive implementation success that leads to system success for a DW implementation. The performances factors are collecting from the literature walk around. These are clustered into the primary drivers of Effective Communication, Training, Education and Documentation, Help, Support and Call Center, Network Management, Software and Hardware source systems, Extract, Transform and Load process and Materialized View.

The drivers, for which metrics can be established and monitored, impact the DW implementation success. Performance success is subdivided into three key components. These include success with Business, Development, and Technological issues that arise during the lifetime of the DW project. Those successes influence the data quality and structure quality, which guide the system success culminating in the supposed success of the overall DW implementation.

The communication process also continues along with the training program. The communication process keeps the business users and IT users in contact with each other to have exchange of views, suggestions and any guidance towards enhanced performance of a data warehouse.

Data warehouse support team members can continue the communication 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 informal parties or get-togethers could be arranged to bridge 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 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.

Network management also plays its part in improving data warehouse performance. From the case study we 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 person(s) 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. 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. The ETL team is headed by an ETL expert. It’s the responsibility of ETL architect to devise a comprehensive and effective ETL process to load the data warehouse. The ETL architect/expert ensures that the ETL processes have strength and endurance.

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.
extraction, data loading, network management, training and communication, query management and some other related tasks. To carry out all these functions and processes a qualified team of full time skilled professionals is required who can efficiently and constantly take care of all the data warehouse maintenance issues in a timely manner. Often the data warehouse projects fail because the users don’t know how to use it according to the business needs. The communication process also continues along with the training program. The communication process keeps the business users and IT users in contact with each other to have exchange of views, suggestions and any guidance towards enhanced performance of a data warehouse.

The help and support an important role in taking valuable output from the data warehouse. Some of the processes like ETL are carried out during the night, which require presence of support staff to rectify any problem. Network management also plays its part in improving data warehouse performance. The hardware and software resources for the data warehouse are compulsory for taking maximum output from it. ETL functions needed to be carried out by a competent and trained ETL team. The ETL team is headed by an ETL expert. It’s the responsibility of ETL architect to devise a comprehensive and effective ETL process to load the data warehouse. The ETL architect/expert ensures that the ETL processes have strength and endurance. The ETL architect works in close coordination with the business users and identifies which data and at what level of detail is required. View materialization is a strategy used to provide fast answers to user queries.

References
[8] Strategy to make superior Data warehouse by Vishal Gour in International Conference on advance computing and creating entrepreneurs Feb2010.

Author Biographies
First Author Vishal Gour received his MCA from IGNOU, New Delhi in 2005. At the present time he is working as Asst. Prof. in the Dept. of Computer Application in Engineering college Bikaner and doing Ph.D in Computer Science from suresh Gyan Vihar University Jaipur, INDIA
Second Author Prof. S.S.Sarangdevot received his MCA and Ph.D in Computer Science from Udaipur, INDIA . His Specialization in Database, Software Engineering & Telecommunications. At the present time he is working as director in the Dept. of Computer Science and Information Technology at JRN Rajasthan Vidyapeeth, Udaipur INDIA