



The Important Features that Improve Data Warehousing Performance

Vishal Goar*
Research Scholar,
Dept. of Computer Science
Suresh Gyan Vihar University
Jaipur, India
goar.vishal@gmail.com

Prof. (Dr.) S.S.Sarangdevot
Director,
Dept. of Computer Science and Information Technology
J.R.N. Rajasthan Vidyapeeth University
Udaipur, India
dr.sarangdevot@gmail.com

Abstract: The purpose of this research is to investigate the important features, which improve the performance of data warehouse technologies in the market. Although a lot of research is going on to enhance the design and development of data warehouses, very little effort has been spent on the maintenance side. Without proper maintenance data warehouse is not going to give the desired output which is expected of it. DW has unique characteristics that may impact the importance of factors that apply to it. In this paper we will try to understand at present available features being used by many companies today to improve the data Warehouse performance. We will try to find the common problems faced. These kinds of problems are faced by data warehouse administrators which are minimizes response time and improves the efficiency of data warehouse overall.

Keywords: Data Warehouse; performance; features; administrators; response time; efficiency

I. INTRODUCTION

Although the term of data warehousing was coined in the early nineties, the global trend is headed for accommodating this technology due to myriad benefits acquired by the adopters. Many reasons have contributed to emergence of data warehousing (as cited previously) in the business field. The lack of convenient awareness, in regard of data warehousing in general and critical success factors in particular, has raised a barrier in front of the adopters [1].

A data warehouse is one of the fundamentals of the decision support systems that are used to support the decision making initiatives, of many IS technologies. Data warehouses have become one of the most talked about information technologies for today's business.

Many Data Warehouse implementations are never monitored for their success. Others are only monitored based upon their technological facets (e.g. speed with which data can be queried) as opposed to the business facets (what enterprise needs are actually being addressed). The major misconception is to treat the success of the Data Warehouse in the same way as the success of an Operational System. [2]

Data warehouse technology is a very costly, time-consuming and risky project compared with other Information technology initiatives; therefore it is important to have a deeper understanding about the features which affect the performance of data warehouse technologies.

II. PERFORMANCE EFFECT FACTOR

While the implementation of a specific phase of the data warehouse may be completed, but the data warehouse program needs to be continued [3]. Progress monitoring needed to be continued against the agreed-on success criteria. The data warehouse team must ensure that the existing implementations remain on track and continue to address

the needs of business. Performance issues in data warehousing are centralized around access performance for running queries and incremental loading of snapshot changes from the source systems [4,5,6] The following six concepts can be considered for a better performance:

A. Interdepartmental cooperation and communication

A Data ware house system is actually about tightly combine different business functions, so the close co-operation and communication across disparate business functions would be a natural prerequisite in an Data ware house project. Some authors have described the co-ordination and communication between departments as the oil that keeps everything working properly in these contexts. [7,8]

The cooperation between the departments in an organization has a large effect on the smooth flow of the required information and expertise among the departments, which strongly influences the successful adoption of data warehouse technology. Interdepartmental cooperation and communication is a must for ant project. 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 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.

B. Education and Training

Training and education of the employees are required in a successful data warehouse project. A data warehouse is not a simple project or an easy-to-learn system. It demands time to

educate and transfer the knowledge to users by setting up training courses and distributing related-material.

Training and updating the employees' knowledge of data ware house is a major challenge. data ware house implementation requires a huge mass of knowledge to enable people to use, cope and solve problems within the framework of the system. Training employees to use ERP is not as simple as training them in any other packaged-software such as a Microsoft package. An ERP system is extremely complex and demands intensive training; it is difficult for the trainers to pass the knowledge to the users within a short period of time. Top management should understand this aspect and should be willing to spend adequate money on educating and training the end users. [9]

C. Vendor Support

In the case of data warehouses, the expensive and the risky nature of data warehouses have forced the potential adopters to pay extra attention in selecting appropriate vendors to increase the possibility of having successful data warehouse initiatives. While training reduces the number of data warehouse questions, a support infrastructure is the key to handling other support needs. [10] In order to ensure success one needs to develop a support structure and plan an approach. When people are using the system, the questions will flow [11].

D. Network Management

If there is a heterogeneous group of platforms for the data warehouse implementation, network management is going to be one of the most demanding tasks [6]. Not only are users coming constantly on-line, but users and equipment are invariably moving to new locations. 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 then there should be at least one person in the organization who understands technology.

E. Extract Transform and Load (ETL)

The ETL process is much more than code written to move data. The ETL architect also serves as the central point for understanding the various technical standards that need to be developed if they don't already exist. These might include limits on file size when transmitting data over the company intranet, requirements for passing data through firewalls that exist between internal and external environments, data design standards, standards for usage of logical and physical design tools and configuration management of source code, executables and documentation. The ETL architect must also ensure that the ETL design process is repeatable, documented and put under proper change control.

A key consideration for the ETL architect is to recognize the significant differences that the design and implementation methods for a business intelligence system have from an online transaction processing (OLTP) system approach. One last role for the ETL architect must be to

ensure that the various software tools needed to perform the different types of data processing are properly selected. ETL is one of the most important sets of processes for the sustenance and maintenance of Business Intelligence architecture and strategy [12].

F. View Maintenance

In a data warehousing environment users queries need to be very efficiently and carefully written as some tables of the data warehouse are very huge and queries posted against these tables could take days or weeks to complete. To have an efficient query management system most of the predefined and ad hoc queries should access summary data instead of detailed data [4]. One needs to employ query navigators to redirect base table queries to the aggregate and summary table level and examine which tables and columns were accessed and the number of rows retrieved. Check response time for these queries and any effects these have for e.g. Paging or locking. Break down the predefined queries into smaller queries for processing. Consider doing most of your resource intensive processing away from the current level of detail and try to run large queries during off-peak hours. This will give more processor time to smaller queries and will aid in getting quick results. Try to push query processing up from the client to the application server level. As part of end user workstation design, consider the employment of thin clients, forcing query processing and scheduling up to the server level. In a data warehouse data should be distributed in a hierarchical manner where the most common information has the least amount of distribution and the least common information has the highest level distribution. Minimize the amount of cross network data retrieval and combination of data from different locations. Allow information access users to follow a hierarchical path when searching for data

III. CONCLUSION

Data warehouse technology is a powerful tool to overcome data-related obstacles and enhance decision making initiatives in our highly globalized and competitive market. A data warehouse solution is not only a software package. It is a complex process to establish sophisticated and integrated information systems. The adoption of this technology requires massive capital expenditure, utilizes a certain deal of implementation time and has a very high likelihood of failure. Therefore, many adoption-related factors must be carefully assessed before the real adoption is actualized.

The results from this study revealed that all organizational and project-related features, first is Inter departmental cooperation and communication is very important feature because having a proficient project team may affect largely the smooth progression of the data warehouse adoption project and if they communicate and cooperate properly the output came better. 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 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 desk and problem management play 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. 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. Network management also plays its part in improving data warehouse performance. We concluded that by having a fast and reliable network user queries get a much shorter response time especially in a distributed data warehouse. 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.

IV. REFERENCES

- [1] An investigation of the factors affecting data warehousing success by Roger L. Hayen et. al. in journal of Issues in Information Systems, Volume VIII, No. 2, 2007.
- [2] Developing a Data Warehouse Process that responds to the needs of the Enterprise, Peter R. Welbrock Smith-Hanley Consulting Group Philadelphia, PA
- [3] The data warehouse toolkit. 2nd edition. Ralph Kimball, Margy Ross. 2002 Wiley computer publishing..
- [4] Data Warehouse Management Handbook by Richard Kachur. 2000 Prentice Hall .
- [5] Building, using, and managing the data warehouse. Ramon Barquin, George Zagelow, Katherine hammer, Mark sweiger, George Burch, Dennis Berg, Christopher Heagele, Katherine Glassey-Edholm, David Menninger, Paul Barth, J.D. Welch, Narsim ganti, Herb Edelstein, Bernard Boar, Robert Small. Data warehousing institute series from Prentice Hall.Y.
- [6] Data warehouse, Practical advice from the experts. 1997. Prentice hall by Joyce Bischoff & Ted Alexander
- [7] Akkermans and Helden, Vicious and virtuous cycles in ERP implementation: A case study of interrelations between critical success factors, European journal of information systems, 2002, Vol.11 Iss. 1, p35..
- [8] Nah et al., Critical factors for successful implementation of enterprise systems, Business process management, Bradford: 2001, Vol.7 Iss.3, p285.:
- [9] Bingi et al., Critical issues affecting an ERP implementation, Information systems management, 1999, Vol.16 Issue3.
- [10] Lessons from a successful data warehouse implementation. Dr. John D Porter and john. J Rome. Arizona State University.
- [11] Building a data warehouse for decision support. 1996 Prentice Hall. By Vidette Poe with contributions from Laura L. Reeves.
- [12] Fundamentals of database systems. 4th Edition. Persons international and Addison Wesley. Ramez Elmasri and Shamkant B. Navathe