Journal of Research in Applied sciences. Vol., 1(1): 15-17, 2014
Available online at http://www.jrasjournal.com
ISSN 2148-6662

ORIGINAL ARTICLE

The Importance of Maintenance For Data Warehouse

Umair Rasheed[*], M.Umer Sarwar

Department of Computer Science, Government College University, Faisalabad, Pakistan

[*]***Corresponding author Email:*** umair514@gmail.com

Abstract: The business industry has taken full advantage of the advancement of technology in the last few decades and various systems and procedures have been invented to make every aspect of an organization perform better and faster. Data warehousing is one of these useful innovations which enables an organization to organize and store vast amount of information related to the operations of the company and to retrieve them easily whenever required. But unlike the traditional information systems the data warehouses need much more support and maintenance. The efficiency of the output that a warehouse system gives is dependent on the way it is managed. Without a proper maintenance procedure the data warehouse is unable to give the expected standard of performance. Too much work is being done on design and architecture aspects of data warehouse. On the contrary very little effort has been spent on the maintenance aspect of data warehouse. So, there is the need to highlight the importance of maintenance for data warehouse. The purpose of this paper is to figure out the necessary maintenance techniques that a data warehouse requires after the deployment and identify the key elements that need to be considered during the maintenance.

Keywords: Data warehouse Maintenance, Maintenance Importance, Maintenance Techniques, Data Warehouse Performance.

Introduction

In the present days the constantly altering business market urges the organizations to have ways to access as much amount of information as possible. Most companies nowadays use information technology for their daily operations but the fact persists that in spite of using many powerful notebook and desktop computers and a high speed network the access to data within the organization is extremely tough or nonexistent. After the advent of data warehousing in the 90s it was presumed that it will develop with considerable speed but unluckily that did not prove to be true.

In 1990s large scale businesses started to find the conventional data storage systems increasingly complex and slow to compete with the rising requirement of timely information. As a solution for this crisis the concept of a place to keep information for the completion of strategic management reports or data warehousing was introduced (Inmon, 2005). The key function of data warehouse is collecting data from various technologies and platforms such as spread sheets, VSAM files and IDMS records and placing them with in a common site which uses a common tool for querying. By this method the databases could be kept on a system which is the most appropriate for operating and the information for strategic reporting could be kept on a mutual location which is easily accessible (Kimball & Ross, 2013).

All organizations including IT produce a vast quantity of data concerning its products, customers, sales, staff and services. But mostly this data stays in the operational systems and cannot come in use of the organization. This circumstance is known as "data in jail". Because of it the company has access to only a small section of the stored data which can have significant negative effects on the profits and sales of the company. Data warehouse is a large compilation of decision supporting information which makes the analyst, manager or employee of a company capable of taking quick and effective decisions based on that technology (Orr, 2000).

Data warehousing is a still developing area and contains various problems. The major issue among them is the maintenance of this system which affects the performance of data warehouse. Performance management of data warehouse is

much like to the designing of a data warehouse. In the life cycle of conventional system there are many levels of planning and analysis. But in data warehouse environment a very little freedom is given to the builders of data warehouse and quick assembling of system is required from them that makes the performance management process a difficult task (Holdsworth & Shores, 1996). If you are facing too much performance problem while using the data warehouse then the feasibility of the project becomes questionable. It is essential to keep in mind that the data warehouse success can only be determined when the data is loaded in it and users become able to make effective decisions by obtaining the data from it. Data warehouse performance is mostly a task of type and quantity of data stored in a database and the workload of data loading work load placed on a system (Holdsworth & Shores, 1996).

In data warehousing a lot of time is consumed in the procedure of extracting, cleansing, processing and loading of data. According to experts about eighty percent of the time taken for creating a warehouse is spent in these tasks. The main cause for the failure of data warehousing project is poor maintenance of these systems. So, the data warehouses need a support from experts on the maintenance aspect (Reddy et al., 2009). The creators of data warehouses often discover issues in the operational process which captures the data from various sources. At that time the captured data requires validation before its storage in the warehouse. Once the data is inside the warehouse several inconsistencies can be found in the field of descriptive data. The developer will need to make modifications to the transaction procedure and purchase or develop a technology for cleaning data (Reddy et al., 2009).

The main reasons for storing the data in warehousing systems are that they could be: cleaned up, reported against and transferred to another data store where they could or cleaned up and/or reported against. Maintenance is an important aspect of data warehouses (DWs) and decision support systems (DSSs) where there is insufficient public debate. A person as a system developer want to know about developments that can affect the data warehouse to allow sufficient time to evaluate what is impacted, to make modifications and test these modifications (Greenfield, 1995).

Problem Statement

The projects of data warehousing are comparatively expensive than the typical operational systems. Therefore, it is extremely necessary for a data warehouse to give the expected performance and results when operating in an organization. Too much work is being done on design and architecture aspects of data warehouse but nobody is considering that what will happen after its deployment. Whether a data warehouse needs maintenance or not? If yes then how to maintain it? So, there is dire need to highlight the maintenance importance and figure out the necessary maintenance techniques that a data warehouse requires after the deployment.

Materials and Methods

A systematic review is a means of evaluating and understanding all available research relevant to a particular phenomenon of interest. A systematic literature review approach is followed in this paper to conduct the review on data warehouse maintenance. The selection criterion through which we assess the study sources is based on the research familiarity of the authors of this work and in order to select these sources we have considered certain limitations: studies included in the chosen sources must be web-available and strictly related to our domain. The review protocol is developed by using the above mentioned keywords and the following list of sources has been considered to conduct the systematic review: ACM digital library, IJDWM, Science Direct, Scopus and Google Scholar. Once the sources had been defined, it was necessary to describe the process and the criteria for study selection and evaluation. The inclusion and exclusion criteria for this study based on the problem statement. Another step in the search process is performed by searching the related work area of the selected papers to boost the review strength by confirming that no valuable reference is missed during the search process.

Results and Discussion

Observing the current business environment the fact is apparent that the information system would constantly be needing developments, advancements and new solutions for its problems in order to keep up with the increasing demands of efficient data organization and storage sources. After the implementation of data warehouse in an organization it should be properly maintained. So, it becomes necessary for an organization to have a fully trained staff that keeps the data warehouse system under constant observation and prevent any oncoming issues before they contaminate the entire system. The fact that the data warehouse systems rely on a vast number of source systems to gather its collection of information and this maximum exposure makes it vulnerable to malfunctions. So, therefore management team of the system will have to possess a thorough knowledge of the different types of changes that occur in the operational systems which can effect negatively on data warehousing systems. However, by a careful and

well thought out maintenance plan that includes meticulous cleaning and sorting of the information stored in the data warehouse on a regular basis and usage of appropriate tools that assist in the purging of the data, the system can be effectively kept from the threat of getting corrupted or too heavily loaded to give the required performance standard. Another important aspect of maintaining a data warehouse is effective training of the users. Many users are assessed to be technology-shy or just reluctant in integrating technology into their daily work which can pose great problems in the usage of this new system. Conducting a training program will help all employees to gain an insightful perspective of the system and make it familiar to them. Moreover, educational programs need to be executed in accordance to the capabilities and skills of the system users. The training programs will work best if divided into different levels which provide learning according to the potential of varying groups. The staff can be trained within the organization by the support team or for advanced training sessions an external party can be contacted. And the advanced training consists of lectures upon modeling of data and functions of extracting, transforming and loading (ETL).

There following are the main elements that should be considered while maintaining data warehouse:

- With the passage of time data warehouse is growing rapidly as the demands for data storage and processing growing.
- Capacity estimation, even if it is based on the exact calculations there is still a chance for conservation which demands from you to consider the early expansion of data warehouse.
- Data staging or temporary data storing is become a big challenge when the source systems are constantly in changing state.
- Advancement in the technology in the form of software, hardware and network requires a quick release changes to be enforced.
- Ad-hoc query access is growing with the time therefore queries should be mentioned carefully because the new users are continually making requests against base table instead of summary tables.

Conclusion

Data warehousing is becoming an extremely convenient technology for data integration and information analysis and being used by business firms for the purpose of effective decision making. But there is a need to give a serious consideration to the maintenance aspect of data warehouse. The most dominating reason for the failure of the project is the factor of negligence of maintenance. It is almost impossible to get the expected outcomes from the system without appropriate maintenance. So, the maintenance is very vital for a data warehouse after its deployment.

References

Brereton P, Kitchenham BA, Budgen D, Turner M, Khalil M, 2007. Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software. 80: 571-583.

Greenfield L, 1995. Maintenance Issues for Data Warehousing Systems. Retrieved August 25, 2014, available at http://www.dwinfocenter.org/maintain.html.

Holdsworth A, Shores R, 1996. Data Warehouse Performance Management Techniques. Oracle Services White Paper. NY: John wiley & Sons.

Inmon WH, 2005. Building the data warehouse. NY: John wiley & Sons.

Kimball R, Ross M, 2013. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. NY: John wiley & sons.

Orr, K. (2000). Data Warehousing Technology (White Paper). available at http://kenorr.com/pg%2033%20d.w.%20 whitepaper.htm.

Reddy SSS, Lavanya A, Khanna V, Reddy L, 2009. Research Issues on Data Warehouse Maintenance. Paper presented at the Advanced Computer Control International Conference on, ICACC'09.