Identification of Useful Maintenance Strategies for Tuning the Performance of Enterprise and Virtual Data Warehouses in Pakistan

Umair Rasheed

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

Email: umair514@gmail.com

ABSTRACT

Data warehousing technology has been used for many years by different organizations for the purpose of planning and decision making. But performance is one of the major issues that the organizations encounter with data warehouse. Maintenance is one of the methods of tuning the data warehouse performance. Different maintenance strategies are available for tuning the data warehouse performance but there is need to identify, which maintenance strategies are useful. The purpose of this study was to identify the useful maintenance strategies that the organizations in Pakistan are using for tuning the performance of enterprise and virtual data warehouses. A self-administered questionnaire composed of 11 questions was used to amass the data regarding data warehouse maintenance from the organizations using enterprise data warehouse and from the organizations using virtual data warehouse. Sampling technique used for the inclusion of respondents was convenience sampling. The collected data were analyzed for descriptive statistics through SPSS software to get the meaningful findings. Recommendations for further research have also been given.

Keywords: *Data warehouse performance tuning, Enterprise data warehouse, Virtual data warehouse, Maintenance strategies*

1. INTRODUCTION

According to [1] the concept of a data warehouse can be defined as source one time and deliver many times. This means that:

- Collect the operational or transactional data.
- Convert the data into a common form.
- Store this data in a central repository which is different from operational system.
- Deliver this data according to requirements of analytical communities from this repository.

A data warehouse is a replica of transaction information mainly structured for querying and reporting. It is an enormous database that stores huge volume of historical data for the organization. The database can be of any type. So, it means it can be a relational database, multi dimensional database, hierarchical database or flat file [2]. A virtual data warehouse (VDW) uses the middleware as a data hub allowing the access to the corporate data stored in various data sources. This virtual approach requires less time and expense to develop, and involves less risk of data being lost whereas an enterprise data warehouse (EDW) is a unified database contains all business related information and accessible all across the company [3]. The executive of companies need the information to make suitable decisions in order to keep the company competitive. They require the information to create the business strategies, set up business goals and objectives. To make strategic decisions the managers and executive must have be adroit to see the data from different business perspectives. So, the data warehouse is the feasible solution of this [4].

The perception of data warehouse was in the trade since the mid 80's but from the start 90's the real significance of data warehouse was recognized. After that each of global 2000 organization attained some sort of data warehousing technology. Since the initial evolution stages of data warehouse many industrial professionals were assuming that this technology will progress rapidly but that not happened. Data warehouse market did not achieve very much since its evolution. Data warehouse users are still complaining about the problems of metadata management, data quality and performance [5]. The data warehouse project can fails in many ways. The data warehouse project can be expensive, important functionalities might not be implemented, the schedule may possibly slip, the users can despondent and the performance can be undesirable [6].

1.1 Problem Statement

Performance is one of the major issues that the organizations encounter with data warehouse after its deployment. Poor performance can twist the successful data warehousing project into the failure and can pose substantial negative effects on the overall performance of a company. There are various methods for data warehouse performance tuning and tuning the data warehouse performance by the means of maintenance is one those methods. Different maintenance strategies are available for the data warehouse performance tuning but which strategies are useful is a big question. So, there is a need to identify the useful maintenance strategies for data warehouse performance tuning.

1.2 Research Question

Which maintenance strategies are useful for tuning the performance of enterprise and virtual data warehouses in Pakistan?

2. METHODOLOGY

In this study the survey research method has been used. The target population was the organizations of Pakistan using enterprise data warehouse and the organizations of Pakistan using virtual data warehouse. Sampling technique used to select the respondents from population was convenience sampling. A self-administered questionnaire composed of eleven (11) questions was used to obtain the responses of the organizations. The questionnaire was delivered to thirty (30) organizations. Out of thirty (30) twenty five (25) organizations recorded their responses in which fifteen (15) organizations were using the enterprise data warehouse and ten (10) organizations were using virtual data warehouse. The collected data were analyzed by the means of SPSS software for descriptive statistics to get meaningful findings.

3. FINDINGS AND DISCUSSION

The findings obtained after survey data analysis are the useful maintenance strategies that are being used by the majority of organizations in Pakistan for tuning the performance of enterprise data warehouse and virtual data warehouse. The detail of these strategies is as follows:

3.1 Use of Specific Persons for the Maintenance Purpose

For the maintenance of both virtual and enterprise data warehouses the organizations in Pakistan are using the specific persons. The organizations with enterprise data warehouse are using more than fifteen (15) persons for the maintenance purpose. On the contrary the organizations with virtual data warehouse are using between 1 to 5 persons for the maintenance purpose.

3.2 Training of Data Warehouse Users

Organizations of Pakistan are using the training methods for data warehouses users. The organizations which are using enterprise data warehouse are giving the training to their users by different methods in which two methods are widely used, which are: arrange training classes for users and training from data warehouse manuals. On the contrary the organizations which are using virtual data warehouse are giving the training to their users by a single method, which is: training from local experts in which the data warehouse experts in the organizations are responsible for the training of users.

3.3 Use of Help desk

Helpdesk is a resource designed to provide users the information about company's products. In Pakistan to solve the issues that the data warehouse users face while using or accessing the data warehouse, the organizations with enterprise data warehouse and the organizations with virtual data warehouse are using the helpdesk. The helpdesk receives the queries from data warehouse users and answers to these queries.

3.4 Network Monitoring and Management

In Pakistan for network monitoring and management the organizations with enterprise data warehouse are using specialized networking persons and networking tools in order to ensure the timely and constant availability of data warehouse to its users. On the other hand the organizations with virtual data warehouse are using only specialized networking persons for network monitoring and management.

3.5 Capacity Planning

Capacity planning is the planning of resources such as hardware and software planning. In Pakistan the organizations with enterprise data warehouse are using the analytical approach for capacity planning in which they estimate the cost of before purchasing the equipments. On the contrary the organization with virtual data warehouse are using the calibrated extrapolation approach for capacity planning in which they track the growth of data warehouse for some time period then estimate the future capacity needs.

3.6 Solving the Issues of Hardware and Software

Solving the issues of hardware and software is an important part of maintenance. In Pakistan the organizations in which enterprise data warehouse has been used, the issues of hardware and software are solved by both vendor from where they purchased the data warehouse and by the experts of organization. On the other side the organizations in which virtual data warehouse has been used the issues related to hardware and software are solved by only experts of organization.

3.7 Extract, Transform and Load (ETL) Monitoring and Management

The ETL process monitoring and management is also very important during the data warehouse maintenance. In Pakistan for ETL process monitoring and management, the organizations with enterprise data warehouse and the organizations with virtual data warehouse are using the ETL team and ETL tools.

3.8 View Materialization

A materialized view gives the indirect access to table by storing the query result. In Pakistan the organizations with enterprise data warehouse are not using the any view materialization approach. On the contrary the organizations with virtual data warehouse are using the view materialization approach and they are using the periodic approach for the maintenance (updating and refreshing) of materialized views. In the periodic approach the views are updated or refreshed after a specific time period such as once in a day or week.

Table 1 and Table 2 represent the maintenance strategies that are being used by the majority of respondent organizations for tuning the performance of enterprise and virtual data warehouses in Pakistan.

Maintenance Strategies	Respondent Organizations in Percentages
Use of specific persons for maintenance purpose	100%
Use more than 15 persons for the maintenance	80%
Use multiple training methods for data warehouse users	87%
Use of helpdesk	87%
Use of specialized networking persons & tools for network monitoring and management	93%
Capacity planning through analytical approach	93%
Experts and vendors are responsible to solve hardware and software related issues	87%
ETL monitoring and management by the means of ETL tool and ETL team	100%
No view materialization	100%

Table 1: Strategies Used by the Majority of Respondent Organizations for Tuning the Performance of Enterprise Data Warehouse

Maintenance Strategies	Respondent Organizations in Percentages
Use of specific persons for maintenance purpose	100%
Use between 1 to 5 persons for the maintenance	100%
Train the data warehouse users from local experts	80%
Use of helpdesk	80%
Use of specialized networking persons for network monitoring and management	80%
Capacity planning through calibrated extrapolation approach	100%

Organization's experts are responsible to solve hardware and software related issues	100%
ETL monitoring and management by the means of ETL tool and ETL team	100%
Materializing the views and use periodic approach for view maintenance (updating & refreshing)	100%

Table 2: Strategies Used by the Majority of Respondent Organizations for Tuning the Performance of Virtual Data Warehouse.

The findings obtained by the means of analysis that was done on the responses of respondent organizations represent the target population.

4. CONCLUSION

Due to the increasing competition in market effective decision making has become very necessary. Therefore, the organizations are heavily relying on technology of data warehousing for the purpose of planning and effective decision making in order to acquire a competitive dominance over the others. But organizations often face the issue of performance with data warehouse that affects the decision making process. So, data warehouse performance tuning is very essential to get desired output from it and to make the decision making process smooth. This research was conducted to identify the maintenance strategies that are useful for data warehouse performance tuning. The strategies identified for two types of data warehouse: virtual data warehouse and enterprise data warehouse. These strategies are: use of specific persons for the maintenance purpose, use of training methods for data warehouse users, use of help desk, network monitoring and management, capacity planning, solving the issues of hardware and software, extract, transform and load (ETL) process monitoring and management and view materialization.

5. RECOMMENDATIONS FOR FUTURE WORK

During this research study I found few areas within the field of data warehousing that can be considered for future research. Firstly, in Pakistan besides enterprise and virtual data warehouses another type of data warehouse is being used by the organizations is "data mart" so, the maintenance of data marts could be considered for the future research. Secondly, some organizations in Pakistan are thinking about outsourcing the maintenance of data warehouses. The advantages and disadvantages of outsourcing the data warehouse maintenance could also be the considered for future research.

REFERENCES

[1]. Oracle, "*Evolving the Data Warehouse: The Next Generation for Financial Services Institutions* (White Paper).", ed. Oracle World Headquarters, USA: Oracle Corporation, 2011.

[2]. R. Mathews, "Simple Strategies to Improve Data Warehouse Performance," Masters Thesis, 2004.

[3]. P. V. Virparia, S. H. Buch, and R. F. Parabia, "Trade and Tricks: Traditional vs. Virtual Data Warehouse," *International Journal of Advanced Engineering & Application,* pp. 220-224, 2010.

[4]. P. Ponniah, *Data Warehousing Fundamentals*: Wiley India Pvt. Limited, 2006.

[5]. S. S. S. Reddy, A. Lavanya, V. Khanna, and L. Reddy, "Research Issues on Data Warehouse Maintenance," presented at the Advanced Computer Control, 2009. ICACC'09. International Conference on, 2009.

[6]. D. M. Walker, "Data Warehouse Project Management (White Paper).", UK: Data Management & Warehousing, 2008.

[7]. R. Kumar, *Research Methodology: A Step-by-Step Guide for Beginners*: SAGE Publications, 2012.

[8]. J. Pallant, *Spss Survival Manual: A step by step guide to data analysis using SPSS*: McGraw-Hill Education, 2010.

[9]. C. M. Thomas and C. E. Begg, *Database Systems: A Practical Approach to Design, Implementation, and Management*: Addison Wesley, 2010.

APPENDIX

1. Which type of data warehouse is being used in your organization?

   a) Enterprise data warehouse

   b) Virtual data warehouse

2. Are there any specific persons in your organization that are only being used for the purpose of maintenance of data warehouse?

   a) Yes

   b) No

3. How many people are involved in your organization for the maintenance or management of data warehouse?

   a) 1 to 5

   b) 6 to 10

   c) 11 to 15

   d) More than 15

4. Is there any training method or mechanism used in your organization for data warehouse users?

   A) Yes

   B) No

   If yes then which training method is followed in your organization? (**Note:** If you organization is using more than one methods then select multiple options)

   a) Training from data warehouse manuals

   b) Arrange training classes for users

   c) Training from local experts

   d) Conferences

   e) College courses

   f) Web based training

5. Is there any helpdesk for data warehouse users in your organization?

   a) Yes

   b) No

   If yes than what is the role or responsibilities of help desk. Specify below:

   _____

6. Which techniques your organization follows for network monitoring or management to ensure that the data warehouse is constantly available to its users?

   a) Use the network monitoring tools

   b) Use the specialized networking persons

   c) a and b both

   d) Other

   If other then specify below:

   _____

7. Which capacity planning (resource planning) approach is followed in your organization for data warehouse?

   a) Analytical approach

   b) The Calibrated Extrapolation approach

   c) Third party approach

   d) Other

   If other then specify below:

   _____

8. How hardware or software related issues are solved in your organization regarding data warehouse?

   a) By vendor from where you purchased data warehouse

   b) By organization itself

   c) Other

   If other then specify below:

   _____

9. How ETL (extract, transform and load) process management and monitoring is done in your organization?

   a) By using ETL monitoring tools

   b) By ETL team

   c) a and b both

   d) Other

   If other then specify below:

   _____

10. Is the view materialization strategy being used?

    a) Yes

    b) No

    If yes then which approach is being used for the maintenance (updating or refreshing) of materialized views?

    a) Lazy approach

    b) Periodic approach

    c) Forced approach

    d) Depend on requirements

    e) Other

If other then specify below:

_____

11. If you are organization is following some additional maintenance strategies for data warehouse or strategies which are not mentioned in questionnaire then specify that strategies below:

_____

_____

AUTHOR's PROFILE

Umair Rasheed received his MS degree in Computer Science from Government College University Faisalabad, Pakistan in 2014. His areas of interest include database and data warehousing. Currently, he is working as Database Developer in Arshad Group of Companies.