

Convert Database Structure into Star Schema Structure for Data Warehouse

Mohammed Abdulameer
Mohammed
Faculty of Information and
Communication Technology
Universiti Teknikal Malaysia
Melaka
76100 Durian Tunggal,
Melaka, Malaysia.
mhmdaldbag@yahoo.com

Prof. Nanna Suryana
Herman
Faculty of Information and
Communication Technology
Universiti Teknikal Malaysia
Melaka
76100 Durian Tunggal,
Melaka, Malaysia.
nsuryana@utem.edu.my

Harith Azam Abdullah
Faculty of Information and
Communication Technology
Universiti Teknikal Malaysia
Melaka
76100 Durian Tunggal,
Melaka, Malaysia.
harith_alyawer@yahoo.com

Mustafa Musa Jaber
Faculty of Information and
Communication Technology
Universiti Teknikal Malaysia
Melaka
76100 Durian Tunggal,
Melaka, Malaysia.
braveheart_vv@yahoo.com

Abstract

The database has started in the 1960s to make designing, building, and maintaining easily for information system difficulties. Since this time the database uses as storage for data and information and solves the problem about saving them safely. The dramatically increase in governments and companies transactions meet by increase in their databases, data storage and queries which used to retrieve data from database. They use information processing system which is used for storage of everyday activities about them. However, information processing systems rely on online transaction processing (OLTP) in DB, which is not so easily accessible to the governments and companies' users. Moreover, relational database was not designed to support multi dimensional view. Need for Multi dimensional view, Online Analytical Processing (OLAP) and reducing time consuming for reports generating leads to the concept of a data warehouse. This study convert database into data warehouse based on a star schema structure by using several tools and techniques as software and hardware. We investigate how star schema makes fast respond for query and for better performance. The star schema structure data can be viewed and analyzed as multi dimensional view and can be used for Online Analytical Processing.

Keywords

Database, Data warehouse, Star Schema

I. BACKGROUND

Decision Support Systems (DSS) are rapidly becoming essential to achieving a competitive advantage for businesses. DSS allows businesses to obtain a huge amount of data that are locked in operational databases and other sources of data and turn data into useful information. Many companies have already construct or are constructing decision-support databases called data

warehouses, in which users can perform their data analysis. A typical data warehouse extracts, integrates and manages the relevant information from multiple, independent, heterogeneous data sources into one centralized repository of information to support decision-making needs of knowledge workers and decision makers in the form of Online Analytical Processing (OLAP) [1].

The data warehouse serves different needs and therefore requires a different approach to design. The concept of a data warehouse is critical, when all data will be stored in the same location and / or method. For decision support, data integrated with multiple sources that can be stored in different database platforms within the different data models or have different data types. Building a data warehouse to display data from different sources, cleaning and translate

Data into a common model populate the database and maintain data warehouses. The creating of data warehouse process consists of five stages: pre-development activities, architecture selection, creation schemes, and population warehouse and data storage services. Each of these steps depends on the previous step [2].

Although the base of operations maintain current information in a data warehouse typically stores information from a historical point of view. Consequently, data warehousing is as a rule. Very large and grow over time. In addition, DSS, users are more interested in identifying the latent form, rather than looking at individual records in isolation. As a result, requests for support decision-making are more complex than online transaction processing (OLTP) queries, and calls for intensive use of units [3].

II. Relative Work

The main problem is the database has limited storage for saving the data that means the database won't be enough for governments or companies' data. That leads to keep less information and weak in analyze because of database features are imperfect as well. Perfect in analyze and huge storage also fast responds for quires and better performance. To meet these needs, there is a new kind of technique that has evolved and matured in the last two decade.

a. Federated database to Data warehouse

Conceptual Federated DW architecture is supporting the tightly coupled integration of independent DMs into a global DW schema. While the global schema hides data heterogeneity from the users, the local DMs retain data and schema autonomy. The canonical data model of our FDW architecture supports DMs implemented on relational and multi-dimensional physical platforms alike. As such, the architecture is ideal for independent organizations that want to share their DWs. Our current and future work involves the implementation of a prototype demonstrating the viability of our approach. The prototype is implemented using off-the-shelf software as far as possible. Particularly, the Oracle 10g and SQL Server 2005 database systems host the dimension and metadata repositories. Moreover, Java technology is used to implement three additional system components: query parser and query processor for SQL-MDi [4] as well as the DM schema integration tool.

The major challenges to overcome are twofold, namely the optimization of the distributed query plans and the robustness of the system. Firstly, the evaluation of queries over the global schema is the crucial factor for achieving a satisfactory performance in DM federations, as pointed out previously [5]. The prototype will include optimization algorithms for query decomposition and query planning similar to those developed in [6]. Secondly, the transmission of the query results is the second major factor on the usefulness of the FDW approach. In case that one or some of the local DMs do not respond to a query request, the federated system should still evaluate the query result. For this purpose, the prototype will apply an "approximates query answering" technique like in [7].

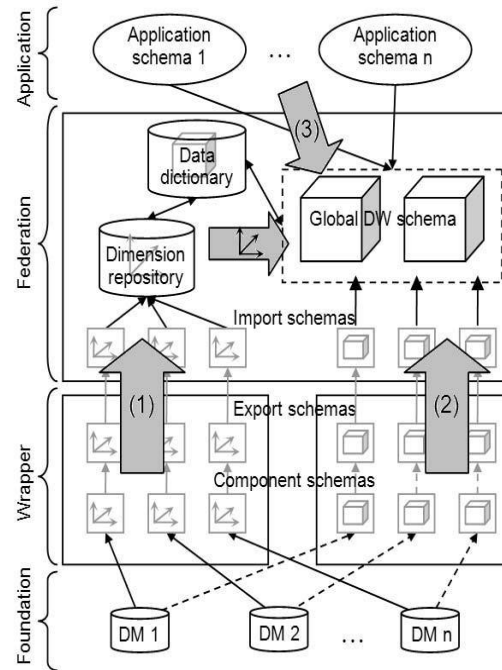


Figure 1. Federated Data Warehouse architecture.

b. Star Schema Structure

The structure of the data warehouse is usually represented by a star schema, facts and dimensions, which are presented in the tables of physical data warehouse. Fact table is located in the centre of the data warehouse and contains foreign keys for all dimension tables [9]. It has many-one with the size of the table. In other words, every dimension record belongs to the thousands of metrics in the fact table. Figure 2 shows the star schema structure.

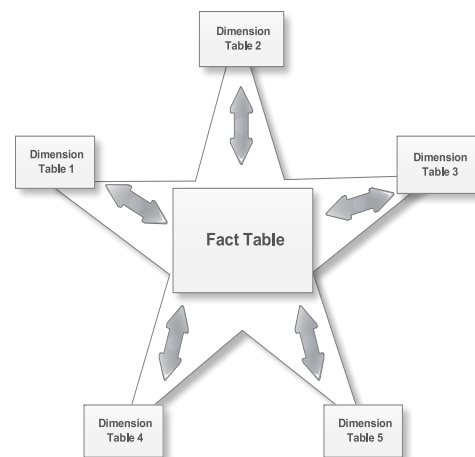


Figure 2 Star Schema Structure

III. Result

This project is concerned; the idea of developed tool for converting operational database to data star schema structure was ignited from the realization of the benefits that other systems will gain from implementing information system specifically a star schema into the conceptual representation of data elements.

This tool developed based on using several software and hardware. In software part employ (programming language and Operating System) so it had used java server pages (JSP), XAMPP (MySql) and Microsoft Windows (Win95/Win98, Windows2000, Windows ME, Windows XP) respectively. In another hand had needs Monitor, CPU, RAM (16MB and above), Disk Space (minimum 12MB).

After log in to the system it will appear the main menu for the converting system that has Generate Star Schema, Drop Star Schema, Change Password, View Generated Star Schema and Logout .The figure 3 shown all of these.

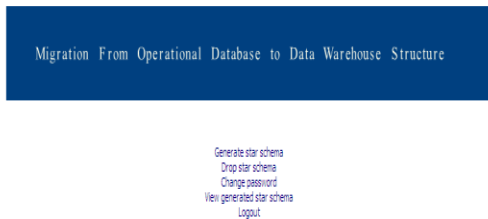


Figure 3: User Page

Figure 4 presents user requires identifying the database source from the Databases and database destination that mean should determine the place where the Star Schema Structure will save in the data warehousing before converting into star schema structure.

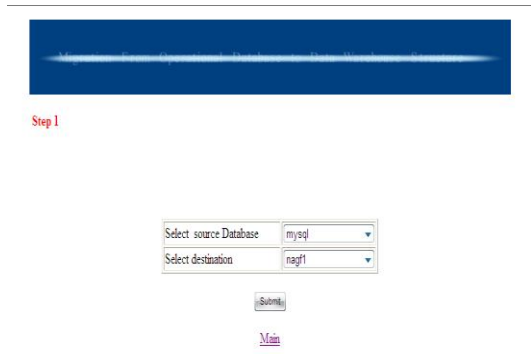


Figure 4: Select database destination page

After identifying all of the database sources, destination, and database table, then users need to identify the name of the Fact table and select the fields from the table that already chose before by the user also select the column from the table depend on the user after write the Fact table name and select the column. In figure 5 shown the page before write the name or select the column.

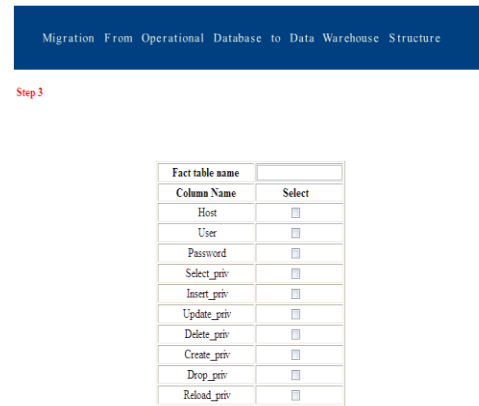


Figure 5: before write Fact table name and Select column

In figure 6 shown the successful page after the user press Submit button to do the converting to the table of database to star schema structure

The star schema generated successfully

Figure 6: Successful page

IV. Conclusion

The database has many issues in design, build, and maintain for information system. Since this time the database uses as storage for data and information and solves the problem about saving them safely. The dramatically increase in governments and companies transactions meet by increase in governments and companies data leads to increase in their data storage and quires.

In these days governments and companies are looking forward to huge repository to save data and information. For that, data warehouse is the best platforms to keep their data. Star schema structures data by create fact table in centre surround by demoralize dimensional tables. This project creates tool by using JSP and MySQL to convert database into star schema structure for data warehouse. Star schema is made fast respond for quire and for better performance in data warehouse.

V. Reference

[1] V. Radiant, (2007). Decision Support System Architecture, Hardware, and Operating System Platforms. Jurnal Sistem Informasi Vol. 2 No. 1.

[2] L. Van, (2008). "A Data Warehouse Model for Micro-Level Decision Making in Higher Education." The Electronic Journal of e-Learning Volume 6 Issue 3 2008, pp. 235 – 244, available online at www.ejel.org.

[3] H. Inmon, (2002). Building the Data Warehouse, 3rd Edition. John Wiley & Sons, Inc., New York, NY, USA, 2002.

[4] S. Berger and M. Schrefl. Analysing Multi-dimensional Data Across Autonomous Data Warehouses. In Tjoa and Tho [26], pp. 120–133.

[5] J. Bernardino, P. Furtado, and H. Madeira. DWS-AQA: A Cost Effective Approach for Very Large Data Warehouses. In M. A. Nascimento, M. T. O' zsu, and O. R. Za'iane, editors, IDEAS, pp. 233–242. IEEE Computer Society, 2002.

[6] M. O. Akinde, M. H. B'ohlen, T. Johnson, L. V. S. Lakshmanan, and D. Srivastava. Efficient OLAP Query Processing in Distributed DataWarehouses. *Inf. Syst.*, 28(1-2): pp. 111– 135, 2003.

[7] S. Berger, and M. Schrefl, (2008). From Federated Databases to a Federated Data Warehouse System.

[8] E. Malinowaki and E., Zimanyi (2006). Hierarchies in a multidimensional model: From conceptual modeling to logical representation.

[9] S. Behrooz, (2003). Teaching Effective Methodologies to Design a Data Warehouse. Retrieved on 25 March 2010, from (<http://proc.isecon.org/2001/35c/ISECON.2001.Seyed-Abbassi.pdf>).