DATA WAREHOUSE IN WEB-ENABLED DATA MINING FOR MOBILE-BASED APPLICATIONS

[1]S.Baskaran, [2]T.SASIREKHA

[1]Head, Department of Computer science, Tamil University (Established by the Govt.of.Tamilnadu), Thanjavur-613010.

[2]Research Scholar, Department of Computer Science, Tamil University, Thanjavur-613010.

ABSTRACT

Construction equipment management and performance data are valuable assets for large contractors that need such historical data for decision making about resource allocation and equipment replacement; however, with large amounts of accumulated data, traditional data analysis based on a transactional system becomes increasingly inefficient. Due to increase in data complexity and manageability issues, data warehousing has attracted a great deal of interest in real life applications especially in business, finance, healthcare and industries. As the importance of retrieving the information from knowledge-base cannot be denied, data warehousing is all about making the information available for decision making. It can help to get better answers which allow both technical and nontechnical users to make much better decisions. Practically, data warehousing and data mining is really useful for any organization which has huge amount of data. Data warehousing and data mining help regular (operational) databases to perform faster. They also help to save millions of dollars and increase the profit, because of the correct decisions made with the help of data mining. A variant dimension of modeling techniques was used to enhance the DW schemas in order to accommodate the requirements of mobile characteristics in the DW design. A proposed mobile DW system was evaluated by expert review, and support the success of mobile DW-based application implementation.

Keywords: ***Data warehouse applications; Decision support systems; OLAP; Preference based***

I. INTRODUCTION

Operational and transactional systems are the new generation systems which are different from 1970's decision support systems (DSS) [1]. In order to complete the life cycle, DSS needs the shadow of a Data Warehouse (DW). A DW pools the available data which is spread all over the organization, and makes a unify pool (like data structure) having the presence of similar and linked formats [2].

Data warehousing takes off in the 1980s as an answer to the very little or no availability of information propagated by online application systems, online applications were praised by a very limited domains of users, and integration was not there even [3]. Historical data kept by online applications are very little as they deposit their historical data for high performance in faster way. Thus organizations hold very little information as compared to data [3].

Mobile phones and smart phones have become very popular for a large number of users (Ketmaneechairat, 2014). Current information and communication technologies (ICT) allow users to get important information instantly. Smart phones are mainstream in this area with active iOS and

Android devices surpassing 700 million globally (Smirnov, Kashevnik, Shilov, Teslya, & Shabaev, 2014). Moreover, mobile technology is rapidly utilized in various computer systems, including Data Warehouse (DW), Business Intelligence (BI) system, and Data Analytic (DA). As reported by Borg and White (2010),

A web-enabled data warehouse can be a viable solution to these problems when considering a number of factors:

1) Data warehousing applications have been used to provide decision support in other industries for over a decade; its subject-oriented and multi-dimensional features fit very well in our application scenario.
2) The general trend of development is the complete separation of the operational system from the analytical system, the former manages Online Transactional Processing (OLTP) while the latter focuses on Online Analytical Processing (OLAP) [1].
3) A data warehousing system, like the Analysis Services in Microsoft SQL Server 2000, is more affordable and provides better cross-platform compatibility.
4) A web-enabled data warehouse has better accessibility, lower costs and requires less of client tools.

The Mtrack Decision Support System (DSS), based on data warehousing technology, has been designed and implemented in this research. The data report and analysis functions in the current Mtrack DSS have replaced those of Mtrack EMS. The paper summarizes our methodology, findings, and the challenges encountered in the design and implementation of this decision support system, with the focus on dimensional modeling and system design.

Adapting the Data Warehouse for the Web

Much is expected of a Web-enabled data warehouse. That means you have to reinvent your data warehouse. You have to carry out a number of tasks to adapt your data warehouse for the Web. Let us consider the specific provisions for Web-enabling your data warehouse. First, let's get back to the discussion of the three stages following the introduction of a new technology. Apart from reducing costsfrom the substitution, demand for data warehouse information has increased.

1) MULTI-DIMENSIONAL MODELING

Based on the current Mtrack Equipment Management System, a data warehouse was planned for the sole purpose of decision support. The objective was to help users gain insight into the equipment data along different dimensional views. Data warehousing is a proven technology for decision support, but it can serve this purpose only if the data warehouse is well planned, designed, implemented and deployed. With this in mind, we identified dimensional modeling and system architecture design as our top priorities. The following paragraphs will focus on the high-level design and modeling of the system, and the technical challenges they posed.

3.1 data warehouse Bus Architecture

Data Warehouse Bus (DWB) Architecture was proposed by Kimball and Ross [5]. So far, it is the most accepted method of data warehouse design. DWB architecture, presented in a bus matrix format, depicts an integrated picture of the whole system and represents a complete set of conformed dimensions and standardized.

DATA WAREHOUSING

To get accurate results from data mining process, current and historical data should be available for the process but keeping the historical data in a regular database would cause a negative effect on the database itself. Usually old data is not used for everyday transactions but it is used for the datamining and reporting issues. Storing historical data in everyday database will cause a huge increase of its size which leads to a slower performance. A good practice is to move the old data from different sources and integrate the whole in another repository called data warehouse [12]. Moving the data from operational databases to a data warehouse involves three steps: 1) cleaning, 2) transformation, and 3) integration [11].

Data warehouse has more than one definitions. The most common one is defined by Bill Inmon who defined it as

the following : "A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process" [1]. As defined , any data warehouse (DW) should have the following characteristics :

Subject-oriented: DW can be used to analyze any subject.
Integrated: DW integrates current and historical data from Different sources.
Time-variant: DW keeps historical data of different time.

Building a web-enabled data warehouse

In 1999, Dr. Ralph Kimball popularized a new term, "data Web house," which included the notion of a Web-enabled data warehouse. He declared that the data warehouse is taking central stage in the Web revolution. He went on to state that this requires restating and adjusting our data warehouse thinking.

In attempting to formulate the principles for building a Web-enabled data warehouse, let us first review the nature of the data Web house. We will use this knowledge to define the implementation considerations. Now let's review the features. Here is a list of the principal features of the data Webhouse:

• It is a fully distributed system. Many independent nodes make up the whole. As Kimball would say, there is no center to the data Webhouse.
• It is a Web-enabled system; it is beyond a client/server system. The distribution of tasks and the arrangement of the components are radically different.
• The Web browser is the key to information delivery. The system delivers the results of requests for information through remote browsers.
• Because of its openness, security is a serious concern.
• The Web supports all data types including textual, numeric, graphical, photographic, audio, video, and more. It therefore follows that the data Web house supports many forms of data.
• The system provides results to information requests within reasonable response times.
• User interface design is of paramount importance for ease of use and for effective publication on the Web. Unlike the interfaces in other configurations, the Web has a definite method to measure the effectiveness of the user interface.

The analysis of the clickstream data tells you how good the interface is.
• By nature, the data Webhouse necessitates a nicely distributed architecture comprising small-scale data marts.

MOBILE DATA WAREHOUSE

The question is whether the DW systems can be modeled and designed from the perspectives of mobile environments need to be explored, and presented the design process for mobile DW systematically. Importantly, the design method will be based on the existing DW modeling (i.e. Dimension Modeling) in order to maintain the consistency of the model throughout the design process.

Moreover, the Dimension Modeling (DM) was proven to be successful used in building DW systems since 1996 (Kimball, 1996; Rizzi, 2007), and most of the new methods are based on the DM paradigm. The concepts of fact, dimension, and measure in DM paradigm will be organized with mobile characteristics in introducing the method for DW design. Generally, any software systems that support mobility is orchestrated into two distinct types of hosts: mobile and fixed hosts Pitoura and Bhargava. Mobile hosts have limited computing power, memory, hard disk, and display screen due to their small size and weight iPhone, Palm Top, Tablet.

The characteristics of the mobile hosts that normally operate in wireless connectivity are identified as low bandwidth, always disconnected, still expensive, small size, small display, limited battery power, easy to lose, and so on. These mobile characteristics should be taken into DW design and implementation. For example, to access and analyze the heterogeneous data sources, disconnected scenario will be aroused. This will cause an error-prone in the results and produced inconsistent information to the users. Instead of resolving on a physical part (e.g., increase wireless signal or bandwidth size), the design of the DW

structure should be able to capture the intermediate results, and finally consolidated or migrated at the available connection. This will be explored and proposed the workable design method to overcome these problems.

Mobile DW made sense in the most specific scenarios such as sales representatives want to promote their business to the prospect client. In the competitive business rival, the sales representatives need to meet client face-to-face and presented the real information at the real-time. With the minimal processing capabilities of the mobile devices (e.g., mobile phone) and prevent to perform deep analytics, a simple set of extracting data from mobile DW is enough to provide the required information. Importantly, the consistency and reliability of the information are controlled by DW structured. A mobile DW is designed to meet business needs that consistent and reliable as accessed by the desktop PC or fixed hosts.

DATA MINING

Data Mining (DM) is a combination of Database and Artificial Intelligent used to extract useful information from huge amount of datasets to help the users to make better decisions. It is usually used as a decision support system [5].

A. Data Mining Usage

Having enormous volume of data, makes it very difficult for human to analyze and get useful information. This causes the importance of using Data Mining techniques. DM is used in different areas to help to extract useful information then make better decisions. For example, DM can be used for marketing purposes. It can help by giving useful information about the best media and time to publish an advertisement which would help to increase the sales of a product. DM techniques (e.g. association analysis) check all the historical related marketing data and compare the sales to provide informative reports to be used by the decision makers then increase the future sales.

CONCLUSIONS

Mtrack data warehouse completely changed the traditional way of analyzing the equipment management and performance data. By using the dimensional modeling and Online Analytical Processing, the system performance data are integrated in a superior structure for reporting and analysis. The analysis of this study shows that the non-governmental organizations use data warehouse technology much more than the government organizations. The governments mostly use data warehouse for controlling the crime and fraud. Non-governmental organizations mostly use DW for data analysis, prediction and making decisions.

REFERENCES

[1] F. Battisti, M. Carli, A. Neri, K. Egiaziarian, A Generalized Fibonacci LSB Data Hiding Technique, 3rd International Conference on Computers and Devices for Communication (CODEC-06), Institute of Radio Physics and Electronics, University of Calcutta, December 18-20, 2006.

[2] C. Shao-Hui, Y. Tian-Hang, G. Hong-Xun, Wen, A variable depth LSB data hiding technique in images, International Conference on Machine Learning and Cybernetics, 2004,, Vol. 7, 26-29 pp.3990 – 3994, 2004.

[3] J. M. Pollard, Theorems on Factorization and Primality Testing, Procedings of Cambridge Philosophy Society, 76 (1974), 521{528.

[4] P. Shor, Algorithms for Quantum Computation:Discrete Logarithms and Factoring , Proceedings of 35th Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press, 1994,124-134.

[5] R. D. Silverman, The Multiple Polynomial Quadratic Sieve , Mathematics of Computation, 48 (1987), 329-339.

[6]. Tutubalin Pavel Innokentievich, Mokshin Vladimir Vasilevich, “The Evaluation of the Cryptographic Strength of Asymmetric Encryption Algorithms “, In the Proceeding of the 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC), pp.180-183, 2017.

[7]. Partha Sarathi Goswami, Tamal Chakraborty, Harekrishna Chatterjee, “A Novel Encryption Technique Using DNA Encoding and Single Qubit Rotations”, International Journal of Computer Sciences and Engineering, Vol.6, Issue.3, pp.364-369, 2018.

[8]. Tausif Anwar, Abhishek Kumar, Sanchita Paul, “DNA Cryptography Based on Symmetric Key Exchange”,

International Journal of Engineering and Technology (IJET), Vol.7, No.3, pp.938-950, Jun-July 2015.

[9] M. Kharbutli, K. Irwin, Y. Solihin, and J. Lee, "Using Prime Numbers for Cache Indexing to Eliminate Conflict Misses," Proc. Int'l Symp. High Performance Computer Architecture, 2004.

[10] V. Krishnan and J. Torrellas, "A Direct-Execution Framework for Fast and Accurate Simulation of Superscalar Processors," Proc. Int'l Conf. Parallel Architectures and Compilation Techniques, Oct. 1998.

[11] D.H. Lawrie and C.R. Vora, "The Prime Memory System for Array Access," IEEE Trans. Computers, vol. 31, no. 5, May 1982.