BUILDING A WEB-ENABLED DATA WAREHOUSE FOR DECISION SUPPORT IN CONSTRUCTION EQUIPMENT MANAGEMENT

Hongqin Fan (hfan@ualberta.ca)
Graduate Research Assistant, University of Alberta, AB, T6G 2E1, Canada

Hyoungkwan Kim (hyoungkwan@ualberta.ca)
Assistant Professor, University of Alberta, AB, T6G 2E1, Canada

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. This paper summarizes a web-enabled data warehouse we developed for a local contractor, as well as the challenges we encountered in modeling and designing such a complex multi-subject Decision Support System (DSS). The methodology proposed in this paper is expected to provide a new approach to data analysis and information delivery in construction management.

Keywords: data warehouse, equipment management, decision support system, web application

1. INTRODUCTION

Standard General Inc. (SG) is one of the major construction contractors in Alberta. With its fast-growing business in equipment rental and road construction, SG developed Mtrack Equipment Management System (EMS) in collaboration with NSERC/Alberta Construction Research Chair in 1998. The new transactional system successfully replaced the traditional, error-prone bookkeeping and reporting processes used for equipment management. Mtrack EMS maintains a parts, fluids and fuel inventory service; stores a historical record of all the servicing performed on each of SG's fleet, which comprises more than 3,000 pieces of equipment; and can produce standardized work orders and more than 48 types of reports on equipment usage, costs, inventory and accounting.

Even though the Mtrack EMS tracks daily operations satisfactorily, the "reporting" functions became increasingly inefficient when large amount of data accumulated over the years of operations. Typical problems such as the following arise: large numbers of reports lack practical flexibility; analysis of some datasets depend on a large number of reports that are restricted with limited query-parameters and a fixed format; in addition, some reports run to as many as several hundred pages. As a result, large quantities of equipment performance and management data, which are valuable assets for the company, are difficult to interpret and prove to be of little value for decision support.

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.

2. LITERATURE REVIEW

A pilot research project and the application of data warehousing technology in construction were conducted by Chau et al. [2] at the University of Hong Kong. The authors built a Decision Support System (DSS) based on Online Analytical Processing (OLAP) for inventory management of construction materials. Based on its successful application to a residential building project, the authors concluded that a DSS that is based on data warehousing can produce more intuitive, multi-view information from the data depository than the traditional OLTP ad-hoc reports. In another case, Berndt et al. [3] reported on the CATCH data warehouse designed for community health decision support: the CATCH system revolutionized resource allocation and health care policy formulation in a comprehensive and systematic manner. The benefits of building the DSS warehouse according to the authors are: (i) Make it possible for local decision makers to organize and interpret the data clearly; (ii) Produce some core reports for local communities at a reasonable cost; (iii) In combination with web, the information can be presented in a variety of formats and distributed more widely in the community. However the authors of these two papers do not adequately address the high-level dimensional modeling and design required for a complex system.

Microsoft also incorporated data warehouse technology into its Enterprise Project Solutions (EPS) [4]. The Portfolio Analyzer in the Project Server allows the participants to analyze the project resources and performance data in the form of multi-dimensional OLAP cubes. It demonstrates that, for complex data structures in project management, a multi-dimensional view of project data provides superior performance over traditional method of data analysis and information delivery.

The Mtrack DSS distinguishes itself from other applications because it focuses on comprehensive subject areas to assess and model all operational processes. The system modeling and design aims to guarantee the final delivery of an integrated, non-volatile data warehouse characterized by easy access and high query efficiency.

In its technical aspects, the system takes advantage of the Microsoft's cutting-edge data warehousing technology, the multi-dimensional data access standard, Office Web Components (OWC) (distributed with Microsoft System Office 2003), and the .NET web development platform. Compared with traditional system implementation and deployment, the approach adopted by Mtrack DSS has a much lower development cost and a significantly shorter development cycle.

3. 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 fact tables. The bus matrix for Mtrack data warehouse is shown in Fig. 1.

1) The rows show the business processes in the system; each process contains interested measurements that serve as System Performance Indicators (SPI) in the fact table.

2) The columns show the conformed dimensions shared by various business processes. Kimball [5] defines the conformed dimensions as either identical or strict mathematical subsets of the most granular, detailed dimension [5].

The construction equipment management system in this research involves complex processes and resources that are no different from those in other construction activities. In this case, system modeling with DWB architecture demonstrates the following advantages:

1) All the processes with interested measurements are identified. The subject-oriented data warehouse is designed to collect, package and present the data accumulated in the business processes. The processes are clearly identified in the rows of the bus matrix.

Common dimensions

Business processes	the_Time	the_Class	the_Department	the_Account	the_Supplier	the_FuelType	the_FluidType	the_Parts	the_Employee	the_Manufacturer	the_Component	the_CostItem	the_RepairCostType
Fuel Consumption	X	X	X	X	X	X				X			
Fuel Inventory	X		X	X	X	X							
Fluid Consumption	X	X	X	X	X		X			X			
Fluid Inventory	X		X	X	X		X						
Parts Consumption	X	X	X	X	X			X		X			
Parts Inventory	X		X	X	X			X					
Purchase Order	X	X	X	X	X			X	X		X		
Work Order	X	X	X							X	X	X	
Repair Cost	X	X	X							X			X
Human Resource	X		X	X					X				

Fig. 1: Bus Matrix for Mtrack Data Warehouse

2) All the common (conformed) dimensions are identified: these dimensions are shared among various processes, so the dimension design will give due consideration to every process in terms of granularity, levels, properties, etc. If well designed, the shared dimensions will guarantee the data consistency and, most importantly, facilitate the data staging; as a result, the data ETL (Extract, Transform and Loading) efforts can be kept to a minimum.

3) The bus matrix can serve as a design tool, management tool and a communication tool [5].

3.2 Multi-dimensional modeling

In the DWB architecture, each row contains a set of measures for the corresponding business process and the associated common dimensions. The best model for this set of data structure can be modeled in a star schema, which includes a fact table at the center, and has all the connected dimensions arranged around it [5]. It models an interested subject in the system, with all the measurements in the fact table and all the data attributes in the dimension tables. The star-shaped data structure enables us to perform system performance analysis along one or along any combination of dimensions. Questions about the measured data addressing when, where, who, by whom, etc., can be answered after the schema is transformed into a dimensional data cube. Hence proper modeling of each star, with its underlying fact table and dimension tables, produces a complete view on the individual subject. All the stars in the system will collectively provide an integrated view of system performance.

In the Mtrack DSS, 10 stars are designed based on the DWB architecture, to model the different facets of the equipment management system. Fig. 2 shows the Repair Cost star schema, one of the most important subjects addressed by the system. The schema includes one fact table and five dimension tables:

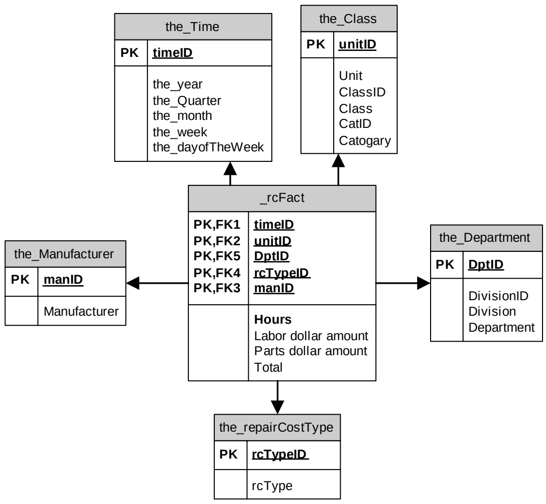

Fig. 2: Dimensional Model for Repair Cost

1) Fact Table: The measures included in the Repair Cost fact table are Hours Spend on Repair, Labor Cost Dollar Amount, Parts Cost Dollar Amount, and Total Repair Cost.

2) Dimension Tables: There are five dimension tables, including Time, Equipment Class, Department, Repair Cost Type, and Manufacturer dimensions. Each dimension has a number of attributes to define hierarchical levels, members at each level, and properties for each member.

All the reports on repair costs in the previous Mtrack Equipment Management System are now condensed into a multi-dimensional data cube: the user can browse, drill down, roll up, or pivot the cost data in the cube along or across any combined dimensions. Examples of questions that can be answered with the system include: show the repair costs of all the earthmoving equipment for the year 2001; compare the repair cost of same class of equipment with different models and manufacturers for a particular period; and show the equipment that has the highest repair cost in a specific class.

During the design of Mtrack Data warehouse, we encountered the following issues that, when dealing with a complex system, must be properly addressed to guarantee the usefulness, efficiency, and non-volatility of the data warehouse:

1) Both the fact table and dimensional tables should focus on the most granular level of daily operations. The data warehouse differs from a database by delivering summarized information rather than a specific transaction; however, the data at the atomic level allows for a higher level view through roll-up, while data modeling at a higher level makes it impossible to drill down to get detailed performance data.

2) All the potentially interested measures in the subject should be included in the fact table. The data warehouse is designed for use by the people at different levels in the company; the more measures are included, the more the diverse needs of users can be satisfied.

3) Some semi-additive measures must be identified: for example, the fuel inventory volume in the fuel inventory fact table cannot be summed up along the time dimension. The cost variation percentage in the Work Order fact table cannot be added up along any dimension, so we choose *average* rather than *addition* for roll-up.

4) The concept hierarchy for each dimension should be properly modeled on and mapped to the dimension attributes or properties. This is the most practical step in order to make best use of measures in the fact table. If possible, all the dimensions shall be organized at different hierarchical levels in a top-down approach; for example, the Equipment Class Dimension has categories, classes and then units. Each unit in the dimension has properties such as the year of make, model, purchase date etc., so that users can compare equipment performance data according to categories, classes, and units, or compare equipment based on the equipment properties.

5) Shared dimensions shall be designed with due considerations to the needs of all the related subjects.

4. DATA WAREHOUSING

Data warehousing refers to all the steps required to build up the data warehouse. According to the dimensional model, these steps include: (i) Identification of data sources; (ii) Data staging, or data Extraction, Transformation and Loading (ETL) from the heterogeneous sources to consolidated data warehouse; (iii) Data presentation to the user via access tools. Fig. 3 shows the processes implemented in Microsoft SQL Server 2000 and the web-based front-end for user interaction with the Mtrack data warehouse.

We chose the Analysis Services in Microsoft SQL Server 2000 to build the data warehouse because of its low cost, high compatibility and ease of use.

4.1 Data Sources

Microsoft Access-based Mtrack Equipment Management System was first put to use in 1998. In 2001, a new SQL Server-based Mtrack EMS began to replace the access-version. This means the historical data sources of the Mtrack database include an Access database and a SQL-server database, their data structures differ slightly from each other.

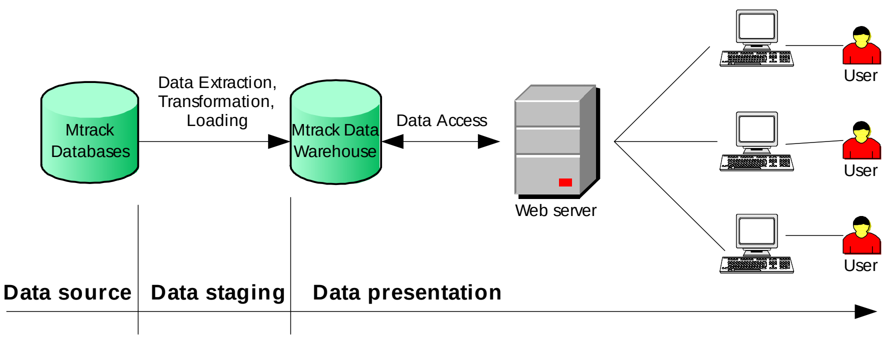

Fig. 3: Data Warehousing and Access

4.2 Data Staging

Data staging means to extract, transform and load the data from data sources to the data staging area. Kimball [5] defines the data staging area as everything between operational data sources and the data presentation area.

In Mtrack DSS, we use the Data Transformation Services (DTS) in Microsoft SQL Server 2000 to implement the data staging process. DTS is a set of visual design tools for data ETL. All the data sources, data validation, extraction and loading of data by SQL scripts, and e-mail notification of execution results can be visually designed as a package comprising a flow of processes. Scheduled execution of the package enables periodic updating of the data in the data warehouse.

A separate database in SQL server is created via the DTS package to provide a staging area where all the required data are cleaned and re-packaged for data presentation.

4.3 Data Presentation

The data presentation area is where the data is organized, stored, and made available for direct query by users, report writers, and other analytical applications [5]. Moving data from the staging area to the presentation area becomes a process of designing, building, and processing the data cubes according to the dimensional models in the data warehousing system.

With a complete set of wizards and visual tools, Analysis Manager in SQL Server Analysis Services makes it fairly easy to build the data cubes, and involves only three consecutive steps: data connection, data cube creation by dimensional model, and cube processing. The star schema for each subject appears in the Analysis Manager exactly as it is in the dimensional model.

After processing the cubes, the user can roll up, drill down, slice, dice, or pivot the data cube with the visual browsing tool in the Analysis Manager; however, to make the data cubes really useful, Multidimensional Expressions (MDX), the query language proposed by Microsoft [4] is needed to define and manipulate the multi-dimensional cubes.

5. MTRACK DECISION SUPPORT SYSTEM DESIGN

5.1 System Architecture

To publish the data cubes to the web and respond to user queries, the Mtrack web application is designed and hosted to work as a bridge between the user and the data warehouse. The technological platform for the communication process is shown in Fig 4. There are three physical parts in the system:

1) Database server: Pivot table service in the Analysis Server of SQL Server 2000 provides a set of client tools for the transfer of dimensional data from the data warehouse. Object Linking and Embedding Database for Online Analytical Processing (OLE DB for OLAP) and ActiveX Data Objects for Multi-dimensional (ADO MD) work as data access adapters for communication with the external applications.

2) Web server: the Mtrack DSS is designed as a web application with Active Server Pages (ASP) .NET development technology; the web server hosts the web application as an interactive web site so that users can query the data.

3) Web browser and Internet connection: the web browser is the only client tool required to communicate with the data warehouse.

5.2 Mtrack Web Applications

Fig. 5 shows a snapshot from the Mtrack web site, there are ten subjects available, each with four modes for data analysis: cube browsing, pre-formulated queries, user-defined queries, and data mining.

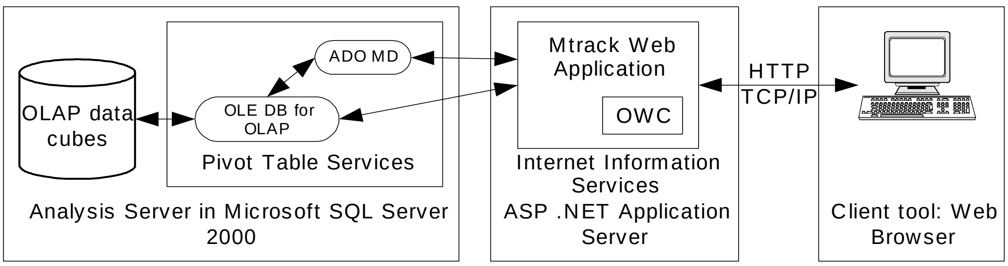

Fig. 4: Mtrack data warehouse system architecture

Microsoft Office Web Components (OWC), involving the pivot table and pivot chart, is embedded in the web pages as ActiveX controls. These ready-use components provide rich interfaces for the visual browsing of data cubes with graphical charts.

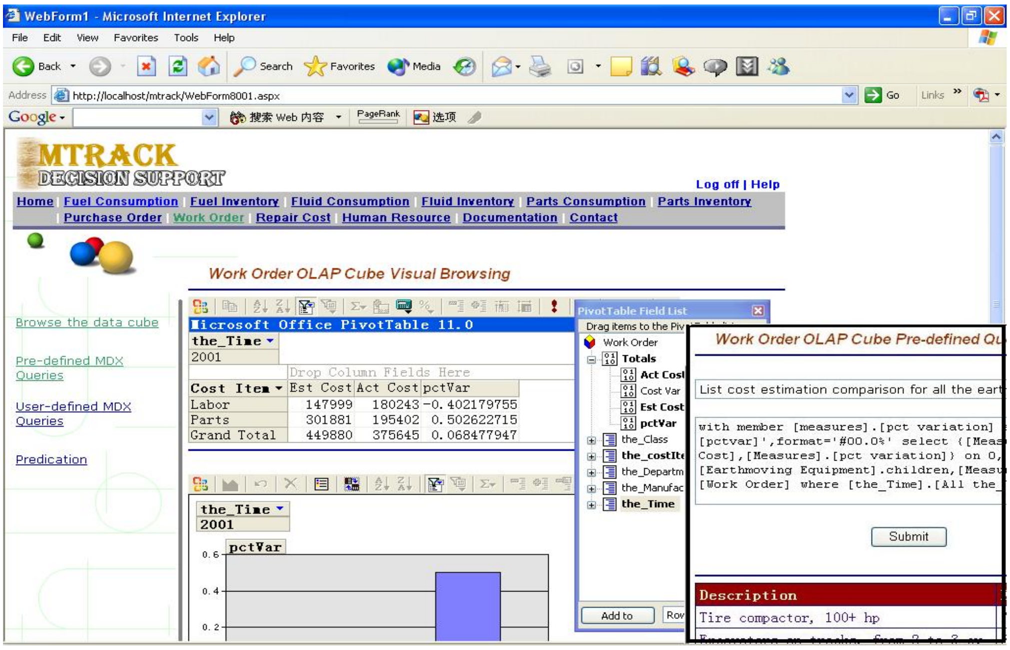

Fig. 5: Web Interfaces of Mtrack Decision Support System

Pre-formulated queries answer the most frequently asked questions on each subject; user-defined queries allow users to define the MDX query, and add it to the pool of queries if authorized to do so. According to Jarke [6], 80~90% of users will depend on cube browsing and pre-formulated queries for data analysis.

The data mining mode is intended for predication and classification based on the historical data. This will be implemented in the next stage of research.

6. 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. This paper presents the dimensional modeling and design of an integrated data warehouse for construction equipment management; a technological platform for deploying the data warehouse to the web is also proposed for quick application development.

Acknowledgements

The authors wish to thank NSERC/Alberta Construction Research Chair for their technical support which made this study possible.

REFERENCES:

[1] T. Bain, M. Benkovich, and R. Dewson. *SQL Server 2000 Data Warehousing with Analysis Services*, Wrox Press Ltd, 2001.

[2] K.W. Chau, Y. Cao, M. Anson, and J. Zhang. *Application of Data Warehouse and Decision Support System in Construction Management*, Automation in Construction, 12 (2002) 213-224.

[3] D. J. Berndt, A. Hevner, and J. Studnicki. *The Catch Data Warehouse: Support for Community Health Care Decision-making*, Decision Support Systems 35 (2003) 367-384.

[4] Microsoft. *MSDN Library online*. http://msdn.microsoft.com/library/default.asp., 2004.

[5] R. Kimball and M. Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, second edition, John Wiley & Sons, Inc, 2002.

[6] M. Jarke, M. Lenzerini, Y. Vassiliou, and P. Vassiliadis, *Fundamentals of Data Warehouses*. Springer-Verlag Berlin Heidelberg, 2000.