Data warehousing for construction equipment management[1]

Hongqin Fan, Hyoungkwan Kim, and Osmar R. Zaïane

Abstract: Equipment logistics, maintenance, and repair are important aspects of construction equipment management. A well-managed equipment fleet helps reduce downtime, as well as total maintenance and repair costs. With quickly growing fleets of equipment, large contractors tend to divert the maintenance and repair of equipment from equipment managers to project managers. As a result, the equipment managers shift their attention from operational-level decision-making to corporate-level strategic decision-making regarding equipment management, which is often a challenging job with the current equipment management system. This paper presents an equipment data warehouse and a prototype decision support system (DSS). The proposed equipment data warehouse enables equipment managers to visually analyze the equipment fleet data from different perspectives and at various level of details. The data-warehouse-based DSS facilitates high-level, fact-based decision-making regarding equipment logistics, supplies, maintenance, repair, and replacement and has higher levels of performance and flexibility than the current equipment management system.

Key words: data warehouse, decision support system, equipment management, multidimensional modeling.

Résumé : La logistique des équipements, la maintenance et la réparation sont des points importants de la gestion des équipements de construction. Une flotte d'équipements bien gérée aide à réduire les durées d'indisponibilité, la maintenance totale et les coûts de réparation. En considérant le rapide accroissement du nombre d'équipements, les grands entrepreneurs ont tendance à faire passer la maintenance et les réparations des équipements d'un gestionnaire d'équipement vers les directeurs de projets. Il en découle que, concernant la gestion des équipements, le gestionnaire d'équipement de l'entrepreneur dirige maintenant son attention de la prise de décision au niveau opérations vers la prise de décisions stratégiques au niveau corporatif, ce qui représente souvent un poste posant un grand défi considérant le système actuel de gestion des équipements. Cet article présente un entrepôt de données sur les équipements ainsi qu'un prototype de système d'aide à la décision. L'entrepôt de données sur les équipements proposé aide les gestionnaires d'équipements d'analyser visuellement les données de la flotte d'équipements en utilisant différents points de vue et divers niveaux de détails. Le système d'aide à la décision basé sur l'entrepôt de données facilite une prise de décision de haut niveau fondée sur les faits quant à la logistique, aux approvisionnements, à la maintenance, aux réparations et au remplacement et démontre un rendement et une flexibilité supérieures par rapport au système actuel de gestion des équipements.

Mots clés : entrepôt de données, système d'aide à la décision, gestion des équipements, modélisation multidimensionnelle.

[Traduit par la Rédaction]

Introduction

In recent years, most large contractors have increased their investment in maintaining, updating, and replacing their equipment fleet to satisfy the needs of project construction (Stewart 2000). Equipment managers in charge of a large fleet need both timely information and tools to make strategic decisions pertaining to resource allocation and equipment maintenance, repair, and replacement. The current trend in the industry is for contractors to redirect the responsibilities of routine equipment maintenance and repair from the equipment manager of the company to project managers (Stewart 2004). The project managers rely on a computerized equipment management system to automate the

Received 4 February 2005. Revision accepted 28 July 2006. Published on the NRC Research Press Web site at http://cjce.nrc.ca on 3 March 2007.

H. Fan. Department of Civil and Environmental Engineering, 3-133 Markin/CNRL Natural Resources Engineering Facility, University of Alberta, Edmonton, AB T6G 2W2, Canada.
H. Kim.[2] School of Civil and Environmental Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea.
O.R. Zaïane. Department of Computing Science, 352 Athabasca Hall, University of Alberta, Edmonton, AB T6G 2E8, Canada.

Written discussion of this article is welcomed and will be received by the Editor until 30 April 2007.

[1]This article is one of a selection of papers published in this Special Issue on Construction Engineering.
[2]Corresponding author (e-mail: hyoungkwan@yonsei.ac.kr).

process of making records on daily equipment operations, preventative maintenance, and repair. However, this system does not provide a substantial benefit to the equipment manager: although most current equipment management systems can provide canned, well-formatted reports, analysis of the equipment data to make decisions turns out to be a nontrivial process.

One of the partners in this research is a major construction contractor in Alberta, Canada. Its current equipment management information system (MIS) was developed in 1997 in collaboration with the NSERC / Alberta Construction Industry Research Chair, partly because of the increase of the contractor's business in road construction and equipment rental. Application of the equipment MIS across the company successfully replaced the traditional, error-prone, paper-based bookkeeping and reporting chores needed to manage the contractor's equipment. By capturing management data on the daily operations of more than 3000 pieces of equipment in the fleet, the equipment MIS maintains a parts, fluids, and fuel inventory service and also stores a historical record of all servicing performed on each piece of equipment. The equipment MIS is also capable of producing standardized work orders and 48 types of reports on equipment usage, maintenance, and repair. Although the equipment MIS tracks daily operations successfully, the equipment manager found that the large amounts of data accumulated over the years made it inefficient to answer simple questions, such as comparing one make of earth-moving equipment with another in terms of total maintenance and repair costs over the last 5 years. The 48 reports in the equipment MIS do provide some degree of flexibility by allowing the user to change a few parameters, such as cost account and time, but the report functions are restricted to a limited number of query parameters and customized formats. Furthermore, some reports are not easy to read, because they can run up to several hundred pages in length. Given the limited number of queries and the fact that the equipment manager must look over the data from a variety of perspectives, these canned, customized reports cannot answer all the constantly changing questions required for decision-making.

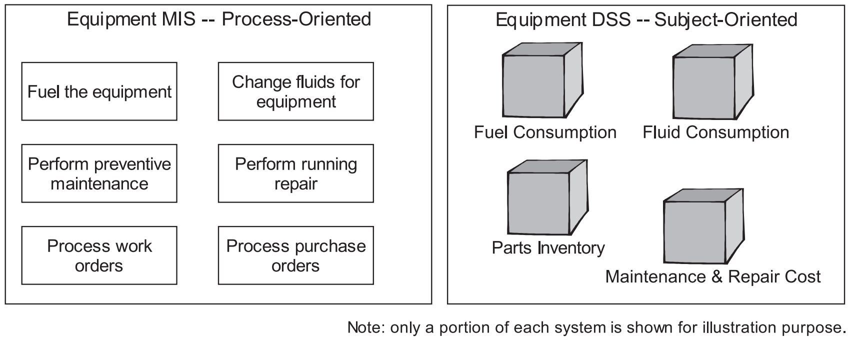
The inability to provide efficient decision support is a well-recognized problem with transaction-processing systems in computer science. Based on a relational database model, a transactional system, such as an equipment MIS, is designed for the efficient capture of operational data. The relational database model is built upon various business processes for daily operations. The equipment MIS mimics purchase order processing, preventative maintenance, fueling, etc., for the collaborating contractor's equipment fleet. The process-oriented equipment MIS guarantees that these data are added and updated efficiently during daily operations; however, sophisticated data analyses are not performed well. The management operations of this process-oriented equipment MIS are structured and repetitive, with isolated transactions. The operations require detailed and up-to-date data, and normally the MIS automates the clerical data-processing tasks. For the sake of decision support, however, queries become very complex and require data consolidated from different sources. Extracting data from a transactional database requires that complex queries be built across different business objects, which can only be accomplished by database specialists. Moreover, extracting data would add considerable unnecessary overload to operational databases. Using an operational system for decision support has become even more inefficient with today's increasing volume and complexity of data. For decision support, when real-time ad hoc complex queries are required for interactive analysis, a multidimensional structure that contains historical data, aggregates relevant information from various sources, and is maintained separately from operational databases is needed. This structure is known as a data warehouse. It allows data to be summarized and analytically processed.

Codd et al. (1993) first tackled this data analysis problem by introducing online analytic processing (OLAP) for decision support. Online analytic processing repackages the data from an online transactional processing system (OLTP) into a data warehouse and presents the data in a multidimensional structure. In addition to its multidimensional nature, a distinct feature of the data warehouse that facilitates dynamic decision support is subject orientation. Subject orientation means that the data model and presentation of the data are centered around individual subjects, such as fuel consumption, parts inventory, and maintenance and repair cost. The fundamental difference between an OLTP-based MIS and an OLAP-based decision support system (DSS) is that the former breaks the system down into managerial functions or business processes, whereas the latter breaks the system down into subjects of interest. Figure 1 compares the process orientation of an equipment MIS with the subject orientation of a DSS. In an equipment data warehouse, data from different OLTP systems are integrated into a single repository and restructured for OLAP. The multidimensional structure of a data warehouse enables data analysis to be performed along any combination of descriptive attributes and at various granularities (levels of detail) for each subject. In brief, the warehouse data come from the original transactional databases but are cleaned up, integrated, and optimized for analyses and reports. The differences between OLAP and OLTP have been described in many references (Codd et al. 1993; Bain et al. 2001).

An equipment DSS using data warehousing technology can provide high-level decision support for equipment managers. In this study, we developed an equipment data warehouse using the historical and current equipment databases of the collaborating contractor. We also designed a Web-based application to expose the data warehouse to an interactive Web site, which would permit remote access and online analysis. Use of this prototype shows that this equipment DSS provides a flexible and powerful environment for equipment managers to analyze equipment data, with clear advantages over the current equipment MIS. This paper summarizes the methodology and findings and the challenges encountered in designing and implementing the equipment DSS, with focuses on the modeling and design of the equipment data warehouse.

The next section is a literature review of related research. This is followed by a description of the data architecture for the proposed equipment data warehouse. The paper then explains the building of the equipment data warehouse and the implementation of the warehouse in a Web-based equipment DSS. The benefits of data warehousing in equipment management are reiterated, and the paper concludes with a summary.

Fig. 1. Comparison between process-oriented view of equipment MIS and subject-oriented view of equipment DSS. Note: DSS, decision support system; MIS, management information system.



Literature review

Chau et al. (2002) conducted a research project and an investigation into the application of data warehousing technology in construction. The authors built a DSS based on OLAP to manage the inventory of construction materials. After successful application in a residential building project, the authors concluded that data warehousing technology could produce more intuitive, multiview information from the data depository than the traditional OLTP ad hoc reports provided. In another case, Ma et al. (2005) applied the data warehousing technique to exchange electronic documents between project participants. In their research project, useful information was extracted from electronic documents and loaded into a data warehouse for an in-depth data analysis by the contractor, the owner, and the resident engineer. The authors concluded that data warehousing enabled all parties to effectively identify critical issues related to their problems.

Microsoft Corporation integrated data warehousing technology into the Portfolio Analyzer of its Microsoft® Office Project Server, which allows project participants to analyze project resources and performance data in the form of multidimensional OLAP cubes. Portfolio Analyzer demonstrates that for complex data structures in project management a multidimensional view of the project data provides superior performance over traditional methods of data analysis and information delivery.

Enterprise resource planning (ERP) is an information technology that integrates a firm's applications and functions into a single computer system with a shared database that is accessible across the enterprise. Enterprise resource planning systems primarily use a relational database. Enterprise resource planning vendors started merging data warehouse and OLAP technology into their product only in recent years. For example, SAP's general platform NetWeaver® (SAP 2006) can be used in the engineering, construction, and operations industry. However, there are few reported cases of ERP systems being implemented by construction contractors; the reasons for reluctance to use ERP systems include the high costs of planning, design, implementation, and training, as well as the delay in return on investment (Ahmed et al. 2003; Shi and Halpin 2003).

Architectural design of equipment data warehouse

The data warehouse provides a consolidated view of enterprise data, allowing users to answer business questions by browsing through the data from different perspectives. To this end, planning and design of the data warehouse need to address two important issues: (*i*) identification of all the relevant subjects from the business processes; and (*ii*) for each subject, identification of a set of measurable facts (numerical measures) and related dimensions (textual entities to describe facts). The dimensions are not designed independently for each subject but are usually shared across subjects. For instance, as shown in Fig. 2, the dimensions "time", "equipment group", and "equipment owner" can all be used for the subjects "fuel consumption", "purchase order", and "maintenance and repair cost". The time dimension is used for all subjects because of the need to evaluate the changes in measures over time.

Kimball and Ross (2002) proposed data warehouse bus (DWB) architecture for the design of a consistent data warehouse: a bus matrix depicting the entire data warehouse is used to identify subjects for operational processes within the company and to obtain a master suite of standardized (conformed) dimensions and facts that are uniformly interpreted across the enterprise. Kimball and Ross defined the conformed dimensions as either identical to or strict mathematical subsets of the most granular (detailed) dimension. We used this approach to design the bus matrix for the equipment data warehouse shown in Fig. 2. The matrix rows show the business processes in equipment management, of which the identified subject represents a data mart or a data cube. Note that these data marts are at the basic level. Data marts at a higher level, if required, can be created by combining the facts from different processes without much effort. Currently, we include 10 basic-level data marts in the warehouse. Each data mart contains the numerical measurements used as equipment management performance indicators. The matrix columns in Fig. 2 show the conformed dimensions that are shared across various equipment management processes. Obviously, the dimensions shared by the most data marts should have the highest priority in the design and deserve the most attention.

Fig. 2. Bus matrix for equipment data warehouse. Note: M&R, maintenance and repair.

Business processes	Time	Equipment Group	Equipment Owner	Account	Supplier	Fuel Type	Fluid Type	Parts	Employee	Manufacturer	Component	Cost Item	Cost Category
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	
M&R cost	X	X	X							X			X
Human Resource	X		X	X					X				

(Common dimensions)

The identification and design of dimensions are critical to determining what kinds of questions can be asked regarding each subject. All the appropriate dimensions should be identified for each data mart before participating in the bus matrix. In most cases, each dimension can be naturally organized in one or more hierarchies, from a higher aggregated level to a lower detailed level, in parent–child relationships; an example is equipment category → equipment class → unit in the equipment group dimension. Therefore, the dimension can be represented as a comprehensive entity organizing similar textual descriptions into hierarchical levels; members at each level can also have their own properties. Depending on the problem domain and how frequently dimensions are shared and used, selection of dimensions may vary across different data warehousing applications. For example, in the equipment data warehouse, the hierarchical structure of the equipment group dimension consists of category, class, and unit in an increasing order of granularity (detail). Although "manufacturer" is a descriptive attribute of the equipment, we chose to build a separate manufacturer dimension rather than use manufacturer as an attribute of equipment dimension, because the manufacturer is commonly shared in the system.

There are many advantages to modeling the equipment data warehouse using DWB architecture. First, all the processes with measures of interest are identified. The processes and data marts are clearly identified in the rows of the bus matrix. Second, all the common dimensions are identified. These dimensions are standardized and shared among various processes, so the dimension design will give due consideration to every process involved. If well designed, the shared dimensions will guarantee the consistent structure and content of a dimension and make it readily usable for different data marts. Third, the conformed dimensions also facilitate data staging by avoiding the repeated pulling of data from the same source. As a result, the data extraction, transformation, and loading (ETL) efforts can be minimized. Finally, a bus matrix can serve as a management and communication tool for a data warehouse (Kimball and Ross 2002).

Multidimensional modeling

In the DWB matrix illustrated in Fig. 2, each row contains a set of measures for the corresponding business process and the associated common dimensions. This data structure can be best represented by a star schema with a fact table at the center and all related dimensions arranged around it. In a data warehouse, the star schema models each subject as a multidimensional data cube, with all the numerical measurements in the central fact table and all the descriptive entities in the surrounding dimension tables. The star-shaped data structure makes it possible to analyze data in the fact table along one or any combination of descriptive dimensions at various granularities. Questions of when, where, who, which, and so on can be answered after the schema is transformed into a multidimensional data cube. Proper modeling of each data cube, with its underlying fact table and dimension tables, enables comprehensive data analysis of an individual subject. All the cubes in the system collectively provide an integrated view of equipment management performance.

In the equipment data warehouse, 10 data cubes were designed on the basis of DWB architecture to model the different facets of equipment management. Figure 3 shows the star schema for the data cube for maintenance and repair cost, one of the most important subjects in the system. The schema includes one fact table and six dimension tables. The fact table for maintenance and repair cost includes measures of the number of hours spent on maintenance and repair, the labor cost in dollars, the parts cost in dollars, and the total maintenance and repair cost in dollars. The six dimension tables are time, equipment group, equipment owner, cost category, account, and manufacturer dimensions.

- The time dimension has two hierarchies: year → quarter → month → day; and year → week → day.
- The equipment group dimension has one hierarchy: equipment category → equipment class → individual unit.
- The equipment owner dimension has one hierarchy: organization owning the equipment → regional division → department.
- The cost category dimension has three categories for cost occurrence: preventive maintenance, running repair, and work order.
- The account dimension contains the financial account descriptions of cost items.
- The manufacturer dimension contains information about the manufacturer of the equipment, including address, contact person, and phone number.

Fig. 3. Dimensional model for maintenance and repair cost data cube.

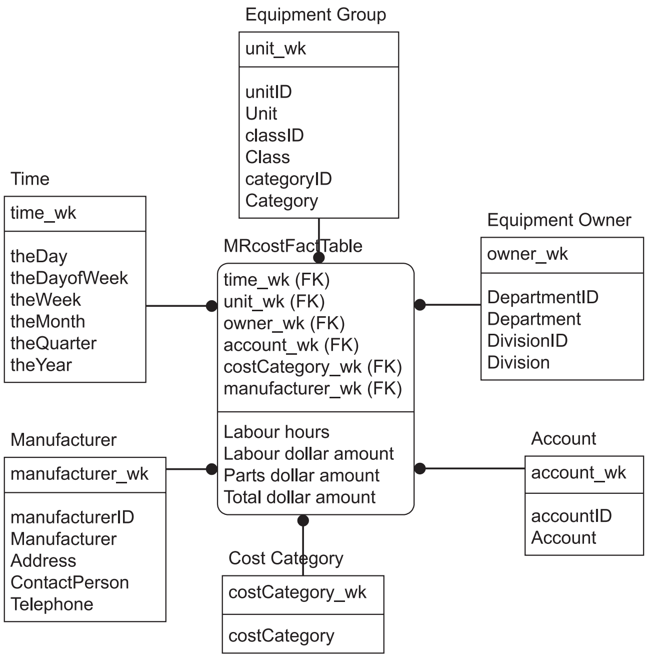
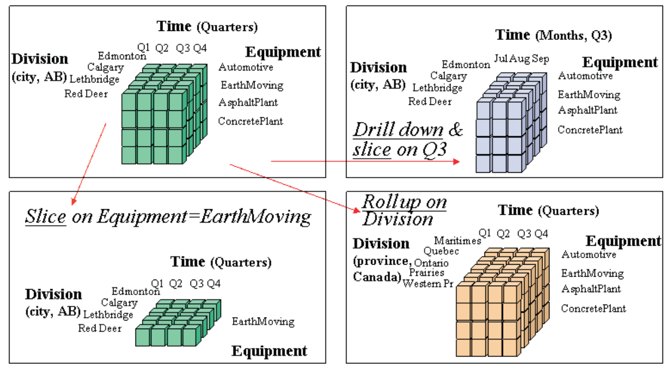


Fig. 4. Examples of OLAP operations on the maintenance and repair cost data cube (drill down and slice, slice, roll up). Note: OLAP, online analytic processing; Q, quarter.



With a multidimensional data cube, all the reports on maintenance and repair costs contained in the previous equipment MIS are now integrated into a single destination. Each data cube can be browsed visually in the data warehouse system. Some important data manipulation and data view functions include drill down, roll up, slice and dice, and pivot, as described below.

- *Drill down* — The drill down operation allows data at a more detailed level to be accessed from a more general category along a dimension. For example, a user could obtain monthly equipment performance data by drilling down through quarterly data.
- *Roll up* — The roll up operation is the opposite of the drill down. These data are manipulated for presentation on a higher level with summarized information along a dimension. For example, monthly data are summarized to get quarterly reports. In the data warehouse system, the numerical facts in data cubes are usually pre-aggregated along each dimension at each level to improve query performance.
- *Slice and dice* — The slice function projects the data in one dimension, whereas the dice function projects the data in two or more dimensions.
- *Pivot* — The pivot function rotates the data cube axes to change the user's view of the data.

With the visual tools in the data warehouse system, the user can browse, drill down, roll up, and slice and dice the cost data in the maintenance and repair cost cube along or across any combination of dimensions, at different levels of detail. Figure 4 shows some examples of OLAP operations on a portion of a maintenance and repair cost data cube containing cost facts along with the dimensions of equipment owner, equipment group, and time (minor changes were made on the dimension members for illustration purposes).

The visual browsing tools are further supplemented by a multidimensional query language that allows the user to get answers to all equipment management questions. The following are examples of multidimensional queries: (*i*) Show the repair costs of all earth-moving equipment for 2001. (*ii*) Compare the repair costs of the same class of equipment with those of different models and manufacturers for a particular period. (*iii*) Show the equipment with the top-*n* repair costs in a particular equipment class.

We encountered several issues that should be addressed when designing an equipment data warehousing system, to guarantee usefulness, efficiency, and consistency.

- *Level of detail in modeling* — The dimensional tables should focus primarily on the most granular level of daily operations. Design of the fact tables should also follow this principle, but not strictly, depending on practical needs and on storage capacity and performance considerations. The data warehouse differs from a database in that it delivers summarized information rather than specific transaction details. The data at the atomic level allow for a higher level view through the roll up operation, while data modeling at a higher level makes it impossible to drill down to get detailed performance data.
- *Measures in the fact table* — All potential measures of interest related to a given subject should be included in its fact table. The data warehouse is designed to be used by different management personnel in the company: the more extensive the measures, the more the diverse needs of users can be satisfied.
- *Rules of aggregation* — Some semi-additive measures must be identified. For example, the parts inventory volume in the parts inventory fact table cannot be summed along the time dimension. The cost variation percentages in the work order fact table cannot be added up along any dimension. Therefore, in these two types of processes, we shall choose *average* rather than *addition* for roll up operations.

Equipment data warehousing

Data warehousing refers to all the processes needed to build up and implement the data warehouse. The procedures for data warehousing include (*i*) identifying data sources, such as operational systems, applications, or flat files;

(*ii*) staging the data, which usually involves data extraction, transformation, and loading (ETL) from heterogeneous sources to a consolidated data warehouse; and (*iii*) using data access tools to present the multidimensional data for reports, interactive analysis, knowledge discovery, etc. Figure 5 shows the equipment data warehousing processes implemented in this research using Analysis Services (a component of Microsoft® SQL Server 2000) and a Web client tool designed for user interaction with the underlying equipment data warehouse.

Data sources

One advantage of using a data warehouse for decision support is that all the data distributed in different systems across the company can be integrated into a single destination. Therefore, all the data sources containing useful equipment management information needs to be identified or the data in the data warehouse will be incomplete and not suitable for strategic decision-making. The data sources in most cases are heterogeneous, which means that they can reside in operational systems (such as in an equipment MIS), applications (such as spreadsheets), and flat text files. The data warehouse system can extract data from these disparate sources and copy them to the data warehouse after preprocessing.

The equipment data warehouse in this research has two data sources: (*i*) the current Microsoft® SQL Server database; and (*ii*) the historical Microsoft® Access™ database used in the first release of the equipment MIS, which contains equipment data from 1997 to 2001. The data accumulated over these years of operations are considered valuable information for the contractor and thus are included in the equipment data warehouse. The current equipment database in SQL Server is used across different divisions located in different areas of Alberta. The data collected from different sources can be replicated and synchronized inside the SQL Server database management system. As a result, only one instance of the SQL Server database needs to be used as the current active data source for the equipment data warehouse. The two sources of the data for the warehouse (the historical Access™ database and the current SQL Server database) have different data structures because of significant changes in the database model of the SQL Server version of the contractor's equipment MIS.

Data staging

Data staging mainly comprises three sequential steps: extraction, transformation, and loading from data sources to the data staging area. Kimball and Ross (2002) defined the data staging area as everything between the operational data sources and the data presentation area. The three tasks of data ETL are completed sequentially to transport data from disparate sources to the data warehouse. The data extraction step copies data from the different sources to a temporary staging database in SQL Server to minimize interference with the operational system. Data transformation is then responsible for preprocessing the data before loading to the warehouse. This step is regarded as being critical to data quality. Potential flaws, such as out-of-range data, erroneous data, duplicate records, and null values, as well as inconsistencies in data format and structure, are resolved in this step. Data transformation is known to be a tedious process, as some deficiencies in the source data are difficult to detect because of the large amounts of data involved. Finally, the transformed data are loaded into the database for presentation. Humphries et al. (1998) indicated that data ETL can consume 60%–80% of the time and effort of the team in data warehousing.

Data ETL is not a one-time process: data in the warehouse need to be synchronized periodically with the updated sources. Therefore, data ETL is usually designed with off-the-shelf software as an application package scheduled to run periodically. For example, we use the Data Transformation Services (DTS) in Microsoft® SQL Server 2000 to implement the data staging process for the equipment data warehouse. Data source definition, data validation, and data extraction and loading, as well as notification of execution results, can be visually designed as a flow of processes in an application package using the complete suite of visual design tools in DTS. Scheduled execution of the package enables periodic updating of the data in the data warehouse.

Data presentation

The data presentation area is where the data are organized, stored, and made available for direct query by users and report writers, as well as by analytical applications (Kimball and Ross 2002). After the data warehouse has been designed, built, and processed, the multidimensional data cubes are stored in an OLAP server for utilization. To expose the data accessible to end users and for interactive analysis, one of the most commonly used approaches is to use OLAP visual browsing tools for interactive data analysis. In this research, we chose to build the visual browsing components into an interactive Web site and host it on a centralized Web server for user access to avoid repetitive installation of OLAP client tools on desktop computers. Warehouse data presentation and access are illustrated in Fig. 5. The next section discusses the design of the Web-based DSS.

Equipment decision support system design

In recent years, many vendors of construction equipment management software have migrated their systems from desktop solutions to Web-based management systems (Computer Guidance Corporation 2005; Caterpillar Inc. 2006) because of the highly distributed nature of construction equipment management operations. The Web-based equipment management DSS proposed in this research aims to provide a mechanism for users to browse and communicate with the equipment data warehouse through a Web site.

Designed as a Web-based data warehouse system (abbreviated Webhouse by Kimball and Merz (2000)), the equipment DSS allows authorized users to remotely access the equipment warehouse with an Internet browser and perform data analysis online. The equipment Web site is developed and works as a bridge between the front-end user and the equipment data warehouse. Figure 6 shows the system architecture, which comprises three logical components.

- *Database server* — In this research project, the Pivot Table Services component in Analysis Services of Microsoft® SQL Server 2000 provides a set of client tools for retrieving multidimensional data from the warehouse. The

Fig. 5. Equipment data warehousing and data access. Note: ADO MD, ActiveX Data Objects (Multidimensional).

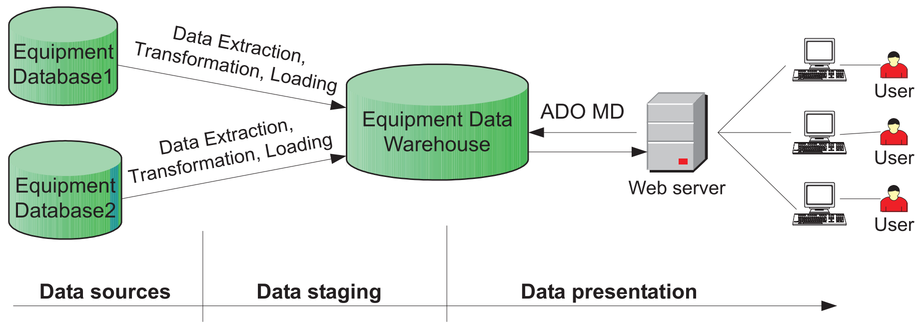
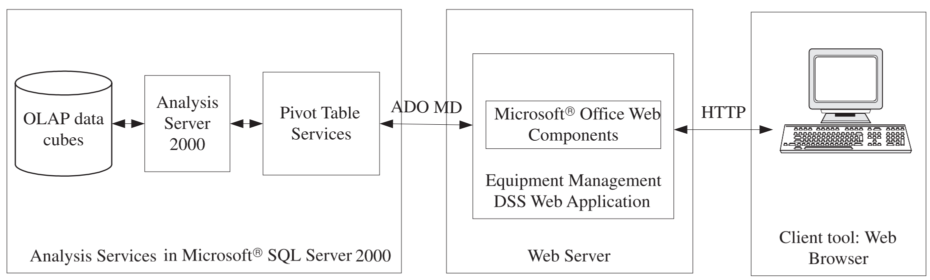


Fig. 6. Equipment data warehouse system architecture. Note: ADO MD, ActiveX Data Objects (Multidimensional); DSS, decision support system; HTTP, hypertext transfer protocol; OLAP, online analytic processing; SQL, structured query language.



Microsoft® ActiveX® Data Objects (Multidimensional) is a data access adapter that enables communication between the data warehouse and external applications.

- *Web server* — The Web application is hosted in the Internet Information Services Web server as an interactive Web site so that users can perform online data analysis after logging in.
- *Web browser and Internet connection* — The Web browser is the only client tool needed for communicating with the equipment data warehouse.

The equipment DSS Web site is an interface that enables users to send in requests and get results from the data warehouse. Therefore, the Web site design should provide a visually intuitive environment that allows users to perform data manipulation and to view data cubes with minimum assistance. In this research, Microsoft® Office Web Components, including Pivot Table and Pivot Chart, were embedded in the Web application to serve as the client-side visual interface for multidimensional data. Pivot Table contains user interfaces for visual browsing of data cubes, and Pivot Chart displays analysis results in a graphical format.

Figures 7 and 8 are snapshots from the equipment DSS Web site. Ten multidimensional data cubes are available in the system. Each data cube has three modes for data analysis: cube browsing, preformulated queries, and user-defined queries. Figure 7 shows a screen that allows the visual browsing of data cubes through roll up, drill down, and slicing and dicing. All these techniques can be accomplished with mouse clicking and the drag-and-drop technique. The measures and dimensions can be expanded from the field list to show members at different levels. Dragging and dropping selected items into different data view windows will slice and dice a data cube. A cross-tabulated table is generated if multiple items are selected for row or column headers in the data view window. Some queries involving special requests, such as top-*n*, sorting, and complex filtering of a data cube, rely on multidimensional query language designed for querying multidimensional databases.

There are no standard query languages for multidimensional databases in the market like the standardized Structured Query Language (SQL) for relational databases; many multidimensional query languages are product dependent. In this research, Multidimensional Expressions (MDX) from Microsoft Corporation (Spofford 2001) is used. The Web-based system provides two approaches for using MDX: preformulated and user-defined queries. The former is predefined in the system to answer the most frequently asked questions. The latter is defined by users on the basis of available templates and can be added to the pool of queries. Figure 7 shows one preformulated MDX query and its running results. According to Kimball and Ross (2002), 80%–90% of all business users will depend on cube browsing and canned queries for data cube analysis. Therefore, predetermining the most important and used queries is paramount.

The impression seems to be prevalent that multidimensional data cube querying is a complex task requiring a high level of expertise. In fact, the MDX query language is greatly simplified compared with SQL queries used against data tables in relational databases. First, the data cubes are read only, and thus, MDX is only used for retrieving data in most cases. Second, each data cube has only one fact table, and dimension tables are equally connected to the central fact table. The simplicity of the star schema structure eliminates the need to join operations from different entities. The purpose of MDX is to define how the user wants to view the subject data — in what perspective and at what level of detail.

Fig. 7. Visual browsing of data cubes in equipment decision support system.

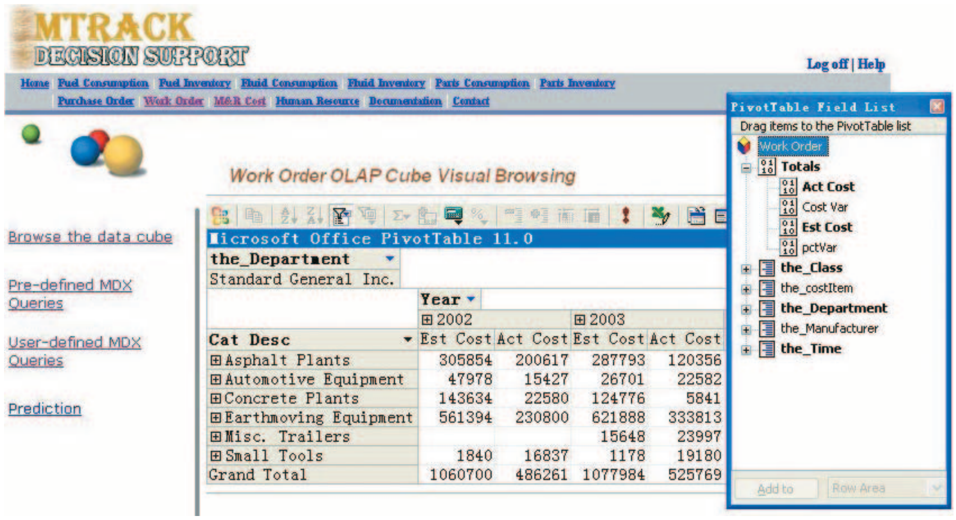
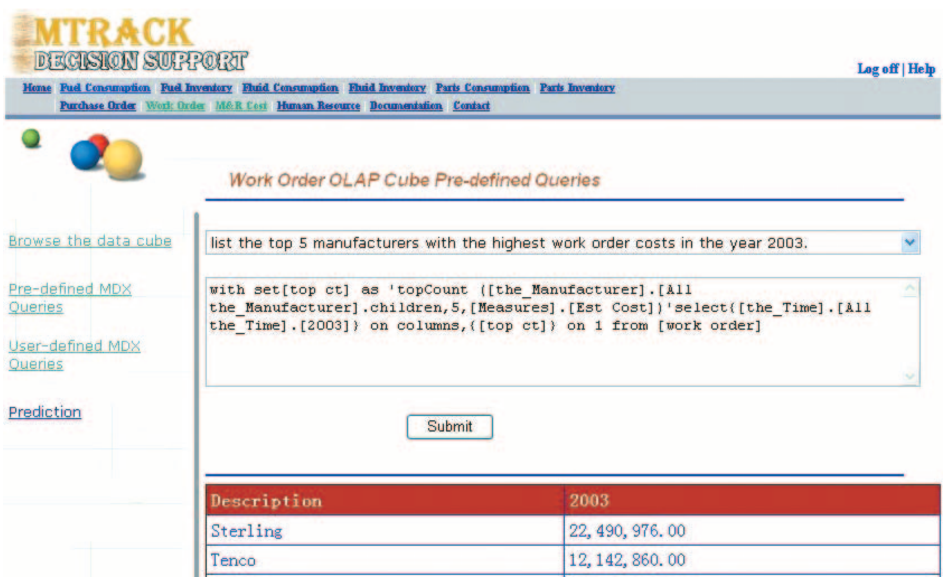


Fig. 8. Customized MDX queries in equipment decision support system. Note: MDX, multidimensional expressions.



The following shows a sample MDX query for comparing the itemized maintenance and repair costs of manufacturer A dump trucks versus those for manufacturer B dump trucks in Division 1 of the contractor's fleet from 2001 to 2004, using the maintenance and repair cost data cube.

SELECT

CROSSJOIN (

{[Manufacturer].[All Manufacturer].[Manufacturer B], [Manufacturer].[All Manufacturer].[Manufacturer A]},

{[costCategory].[All Category].MEMBERS}) **ON COLUMNS**,

{[Time].[All Time].[2001] : [Time].[All Time].[2004]} **ON ROWS**

FROM [MRcostCube]

WHERE ([Equipment Group].[All Equipment Group].[Automotive Equipment].[Dump truck or deck (6 wheels)], [Equipment Owner].[All Equipment Owner].[Divison 1.], [Measures].[Total])

In the example, SELECT is followed by data sources for column and row headers, FROM specifies the data cube, and WHERE defines the filtering criteria. The running results from the system are shown in Table 1. Note that the data in the table have been modified for reasons of confidentiality.

To answer the same question with SQL and the original equipment database, a complex SQL script needs to be designed, which would involve at least a dozen tables (normalized) and complex SQL statements. Parameterized SQL-based reports in the original equipment MIS can answer this question provided this problem is addressed in the system design, but the reports cannot cater to the need to answer unanticipated, constantly changing equipment management questions. Furthermore, this sample question can be answered visually using OLAP-supported visual tools, which automatically gen-

Table 1. Results of sample MDX query on maintenance and repair cost data cube.

	Manufacturer A			Manufacturer B		
	Preventive maintenance ($CAN)	Running repair ($CAN)	Work order ($CAN)	Preventive maintenance ($CAN)	Running repair ($CAN)	Work order ($CAN)
2001	46 794	591 964	91 444	176 154	3 230 614	25 278
2002	86 060	1 190 756	643 683	441 512	3 218 389	1 319 932
2003	78 472	1 193 318	262 688	470 044	1 849 637	2 031 146
2004	92 194	949 521	309 083	561 815	1 547 489	1 828 335

Note: The data in the table have been modified for reasons of confidentiality.

erate the MDX statements behind the scene. Users do not need to know MDX.

Benefits of equipment data warehouse

The developed equipment data warehouse was deployed via a prototype Web-based DSS and tested in a local network. It proved that building and deploying a separate equipment data warehouse based on the current equipment data not only eliminated performance and maintenance interference with the original equipment database, but also provided a more flexible and powerful environment for strategic high-level decision support in equipment management. Major advantages over the current system for equipment data analysis include the following.

- *Better user control of data analysis* — Equipment managers usually query the equipment data in broad and unexpected ways to make high-level strategic decisions. The limited flexibility of the current system in presenting equipment data cannot satisfy the various needs of equipment management because of its inherent characteristics of “easy to get data in, difficult to get data out”. The equipment data warehouse helps equipment managers accomplish these tasks through the roll up, drill down, and slicing and dicing of different multidimensional cubes without additional assistance from database experts. To answer complex business questions, the multidimensional query language for the generation of datasets from multidimensional data cubes is simpler than SQL for relational databases.
- *A better tool for problem identification and investigation* — The body of equipment data collected for a large fleet on a daily basis is huge, and thus potential problems arising from deteriorated equipment or inappropriate decisions at an operational level are not easy to detect with the current equipment MIS, which relies on the concept of a relational database. However, the equipment data warehouse has the capability to help equipment managers gain valuable insight into these data. The numerical data in the fact table, as performance indicators of equipment management, can be aggregated along different dimensions at different levels or properties for analysis. This enables equipment managers to detect problematic areas through interactive exploration of equipment data. For instance, the equipment manager can analyze the labor cost in maintenance and repair works across the company by simply choosing the maintenance and repair cost data cube in the equipment data warehouse and making comparisons over time, according to equipment classes, manufacturers, divisions, departments, and so on. An exceptionally high labor cost in one division for the equipment of a particular manufacturer will be noticeable immediately. The equipment manager may then further investigate the problem by looking at other details. If the system indicates that the equipment from one manufacturer is not being properly maintained or that the mechanics in the division do not have the expertise for this group of equipment, the equipment manger can then decide, for example, whether to outsource the maintenance and repair work to a subcontractor or local dealer.
- *A better tool for strategic decision-making* — Strategic decision-making in equipment management deals with long-term, high-level issues or corporate policies. Strategic decisions answer questions such as “Shall we replace the group of equipment we purchased from this manufacturer in 1985?”; “Shall we outsource the equipment maintenance and repair for this particular project?”; “Which one of two competitive manufacturers shall we buy new equipment from?” Obviously these questions cannot be answered with a few ad hoc reports. Comprehensive exploration of the historical data using an equipment data warehouse can assist with these decisions by providing a consolidated view of equipment management data across the company.
- *Improved data sources for knowledge discovery* — Knowledge discovery from databases helps the user to detect hidden patterns (common or unusual) and trends, as well as to predict events. Data in the equipment data warehouse have improved quality, integrated views, and well-organized structure and therefore can be used as for improved pattern identification and predictive analysis.

Summary and conclusions

Construction equipment management has been drastically simplified because of wide application of equipment management software by large contractors. Large amounts of equipment data collected during daily operations and managements need to be transformed into actionable information for high-level decision support. This paper summarizes a research project in which a separate equipment data warehouse was built upon the equipment databases that were currently being used by a construction contractor. The data warehouse was deployed with a Web server for decision support. In this research project, equipment data from two different equipment databases were consolidated into a single data repository after preprocessing; the equipment data were repackaged into subject-oriented data cubes using multidimensional data models for visual interactive analysis. The equipment data warehouse enables the equipment manager to gain valuable insight into the contractor's equipment data and to answer various equipment management questions in real time. As a result, the decisions on equipment management are based more on facts and less on individual experiences.

The design of the equipment data warehouse was based on bus architecture and multidimensional modeling. The general procedures for designing, implementing, and deploying an equipment data warehouse are introduced in this paper. The challenges and best practices in building an equipment data warehouse are also discussed.

This research showed that data warehousing concepts and techniques can be applied to computer-assisted construction equipment management to improve the current practice of decision support. This was affirmed by industry practitioners after preliminary use of the system. With the rapid development of computer technology, the high cost of designing and implementing a data warehouse has dropped to a reasonable level. The major providers of relational database products have included data warehouse technology in their database products as either built-in features or add-on products. Therefore, building an equipment data warehouse for decision support in construction equipment management is not only feasible and efficient, but also cost effective.

Acknowledgements

This work was supported by Natural Sciences and Engineering Research Council (NSERC) grant No. CRD 226956-99. The writers wish to express their sincere appreciation to Dale Tillapaugh, the equipment manager of the collaborating contractor.

References

Ahmed, S.M., Ahmad, I., Azhar, S., and Mallikarjuna, S. 2003. Implementation of enterprise resource planning systems in the construction industry. *In* Winds of Change: Integration and Innovation of Construction. Proceedings of the ASCE Construction Research Congress, Honolulu, Hawaii, 19–21 March 2003. [CD-ROM.] *Edited by* K.R. Molenaar and P.S. Chinowsky. American Society of Civil Engineers, Reston, Va. doi: 10.1061/ 40671(2003)125.

Bain, T., Benkovich, M., Dewson, R., Ferguson, S., Graves, C., Joubert, T., Lee, D., Skoglund, R., Turley, P., Youness, S., and Scott, M. 2001. Professional SQL Server 2000 data warehousing with Analysis Services. Wrox Press Ltd., Birmingham, Ala.

Caterpillar Inc. 2006. Caterpillar EquipmentManager [online]. Available from http://www.cat.com/cda/layout?m=44501&x=7 [cited 6 June 2006].

Chau, K.W., Cao, Y., Anson, M., and Zhang, J. 2003. Application of data warehouse and decision support system in construction management. Automation in Construction, **12**(2): 213–224.

Codd, E.F., Codd, S.B., and Salley, C.T. 1993. Providing on-line analytical processing to user–analysts: an IT mandate. E.F. Codd and Associates, San José, Calif.

Computer Guidance Corporation. 2005. Construction Management System: eCMS [online]. Computer Guidance Corporation, Scottsdale, Ariz. Available from http://www.computer-guidance.com/products_construction/financial_management.asp [cited 6 June 2006].

Humphries, M., Hawkins, M.W., and Dy, M.C. 1998. Data warehousing: architecture and implementation. Prentice Hall PTR, Upper Saddle River, N.J.

Kimball, R., and Merz, R. 2000. The data Webhouse toolkit: building the Web-enabled data warehouse. John Wiley & Sons, Inc., New York, N.Y.

Kimball, R., and Ross, M. 2002. The data warehouse toolkit: the complete guide to dimensional modeling. 2nd ed. John Wiley & Sons, Inc., New York, N.Y.

Ma, Z., Wong, K.D., Heng, L., and Yang, J. 2005. Utilizing exchanged documents in construction projects for decision support based on data warehousing technique. Automation in Construction, **14**(3): 405–412.

SAP Inc. 2006. SAP NetWeaver® [online]. Available from http://www.sap.com/platform/netweaver/index.epx [cited 6 June 2006].

Shi, J., and Halpin, D.W. 2003. Enterprise resource planning for construction business management. ASCE Journal of Construction Engineering and Management, **129**(2): 214–221.

Spofford, G. 2001. MDX solutions with Microsoft SQL Server Analysis Services. John Wiley & Sons Inc., New York.

Stewart, L. 2000. Giants replace machines to control costs. Construction Equipment, **102**(3): 62.

Stewart, L. 2004. Reliability enlists project support for maintenance. Construction Equipment, **107**(10): 59.