Safety Information Management in Construction Firms: A Data Warehousing Approach

Salman Azhar, Ph.D.
Assistant Professor, Department of Building Science
Auburn University, Alabama, USA

Abstract

Safety statistics for construction indicate high fatality and injury rates. Construction firms in the USA must keep detailed records of all site incidents for OSHA (Occupational Safety and Health Administration). It is found that many construction firms maintain these records just to satisfy the OSHA requirements and seldom use the data in the planning and decision-making of future projects. If the safety related data are integrated from different projects over a period of time, they can provide a good overview of the safety performance of general contractors and subcontractors involved in different trades and under different project conditions. A questionnaire survey by the author indicated that a majority of construction firms in USA use spreadsheets to record and report the site incidents. The data stored in these spreadsheets are non-integrated and often available in non-user friendly forms and formats and hence cannot be readily used by the decision-makers. In this research, a <u>Safe</u>ty Information <u>Man</u>agement System, named *SafeMan*, is developed using the data warehousing technique. Data warehousing is an improved approach for integrating data from multiple heterogeneous databases and other information sources. The *SafeMan* enables users to analyze safety data of past and existing projects to evaluate safety performance of contractors and subcontractors. The paper first discusses the importance of safety information management and basic concepts of data warehousing technique. Next, a step-by-step methodology is presented to illustrate the different stages of the *SafeMan* system development. At the end, discussion is made on a prototype system developed for reporting and analyzing safety information in a construction organization.

Keywords

Safety analysis, Construction site management, Information management systems, Data warehousing

1. Introduction

Safety is one of the primary concerns of the construction industry. The accident rate in construction is reported as the highest in comparison to other industries in USA. There were 1,184 fatal occupational injuries and 421,400 nonfatal injuries and illnesses in construction in the year 2005 (data for year 2006 is not yet available). Incidence rates for nonfatal injuries and illnesses in construction were 6.3 per 100 full-time equivalent workers as compared to 4.6 per 100 full-time equivalent workers in all private industries in 2005 (Bureau of Labor Statistics, 2007). The total cost of occupational injuries/illnesses and fatality incidents can threaten the survival of a construction company in a highly competitive environment. A company operating at a 4% profit margin would have to increase contract prices by $400,000 to pay for a $16,000 injury, such as the amputation of a finger (Ahmed *et al*., 2006).

Construction firms in USA must keep detailed records of all site incidents (minor and major) for OSHA (Occupational Safety and Health Administration) as well as report specific types of illnesses or injuries. If

a firm employs eleven or more employees at any one time in a calendar year, it must file OSHA Forms 300, 300A, and 301. OSHA 300 is the log of work-related injuries and illnesses, OSHA 300A is the summary, and OSHA 301 is a supporting document that is used to record additional information pertaining to each injury and illness logged on OSHA Form 300 (Mincks and Johnston, 2004).

A questionnaire survey conducted by the author indicated that a majority of construction firms in USA use informal spreadsheets to record and report the site incidents (Azhar, 2005). It is found that many construction firms maintain these records just to satisfy the OSHA requirements and seldom use the safety data in the planning and decision-making operations. If the safety related data are integrated from different projects over a period of time, it can provide a good overview of the safety performance of general contractors and subcontractors involved in different trades and under different project conditions. The construction firms can also use these data to improve their safety performance in future projects.

The survey results also indicated that the safety information management systems in use are good for storing and managing the safety records but are not very effective in integrating and analyzing the safety data such as by generating safety data trends in a specific trade over a period of time or by performing scenario analysis. Moreover, the safety data managed via these systems were found to be non-integrated and stored in forms and formats that made it difficult for the decision makers to make quick decisions (Azhar, 2005). Another reason behind the low performance of these systems is the use of traditional *Online Transaction Processing* (OLTP) database technology. The OLTP databases are suitable to answer *who* and *what* type questions, they are not very effective in answering *what-if*, *why*, and *what-next* type queries (Ahmad and Azhar, 2005).

In this research, a <u>Saf</u>ety Information <u>Man</u>agement System, named *SafeMan*, has been developed using the data warehousing technique. Data warehousing is an improved approach for integrating validated data from multiple, often very large, distributed, heterogeneous databases and other information sources. Data warehousing is based on *Online Analytic Processing* (OLAP) concept. The OLAP enables users to analyze multi-dimensional data using different analytical models and tools. The users can experiment with different views of the data; generate data trends over a period of time to make any forecasts; and can perform scenario (e.g. what-if or what-next) analysis.

In the following sections, the concepts of data warehousing, Online Analytic Processing (OLAP) and advantages of a data warehouse over traditional database management systems are reviewed. Then a step-by-step methodology is presented to illustrate the different stages of the *SafeMan* system development. After that, a prototype system developed to manage and analyze safety information for a construction organization in Florida is demonstrated. At the end, discussion is made on the basic functionalities and limitations of the *SafeMan* system.

2. Data Warehousing Concept

Data warehousing is an important development in the field of information systems (IS). A latest study shows that the data warehousing market has grown from $8.8 billion in 1996 to around $200 billion in 2005. Majority of big enterprises are quickly implmenting data warehousing, as evidenced by a report that indicates more than 90% of the large Fortune 500 companies have either adopted or decided to adopt data warehousing (Azhar ahd Ahmad, 2006).

A data warehouse is a dedicated read-only database created by combining data from multiple databases for purposes of analysis. The traditional view of database management is based on the need for data to support transaction processing. Conversely, data warehousing evolved as an answer to the need to support analytical processing for informed decision-making. The primary purpose of a data warehouse is to

provide easy access to specially prepared data that can be used with decision support applications, such as management reporting, queries, data mining and executive information systems.

A data warehouse differs from a traditional database in the following aspects (Ahmad and Azhar, 2005): (1) it covers a much longer time horizon (several years to decades) than do transaction systems; (2) it includes multiple databases that have been processed so that the warehouse's data are subject oriented and defined uniformly (i.e., "clean prearranged data"); (3) it contains non-volatile data (i.e. read-only data) which are updated in planned periodic cycles, not frequently; and (4) it is optimized for answering complex queries from direct users and applications.

Typically, data warehouse architecture has three components or tiers, as shown in Figure 1. Tier 1 (back end) consists of data acquisition tools that extract data from transactional databases and other information sources, consolidate and summarize the data, and load it in to the data warehouse. Tier 2 contains central data repository for managing and storing the data; while Tier 3 (front end) holds the client softwares that enable users to access and analyze the data stored in the warehouse.
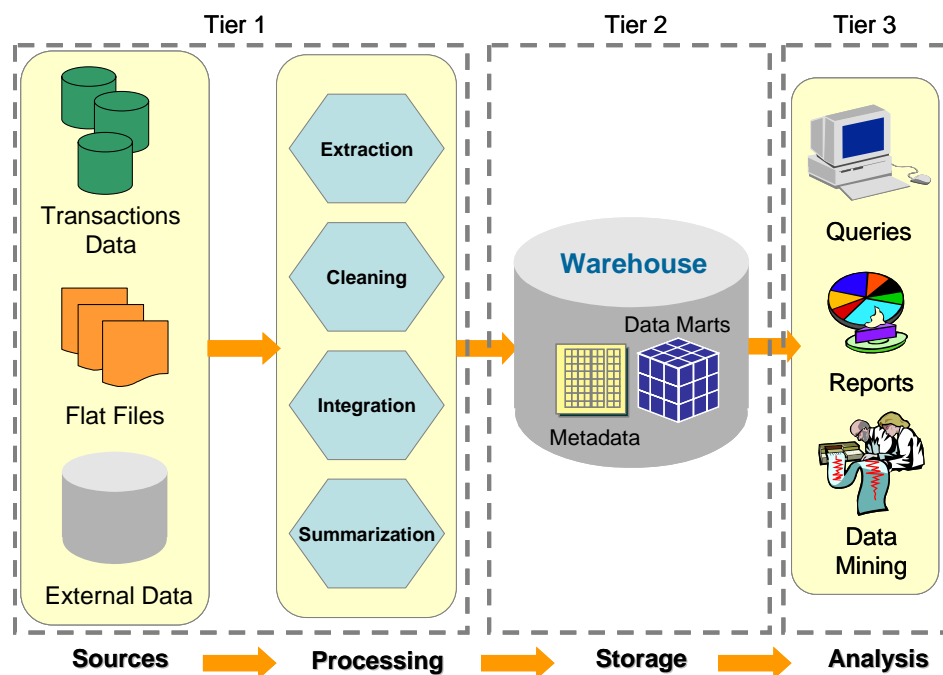

Figure 1: Conceptual Schema of a Data Warehouse

Online Analytic Processing (OLAP) technique is the backbone of a data warehouse. OLAP enables users to analyze data stored in a data warehouse using different analytical models and tools, such as the Analytical Hierarchy Process (AHP), the Expected Utility Value (EUV) based on decision tree models or any other numerical/analytical techniques. With the appropriate user-friendly query tools, one can experiment with different views of the analyzed data, thus gaining a better insight into the situation in order to make informed decisions (Ahmad and Azhar, 2005).

3. Methodology and System Development Framework

A three phase methodology as shown in Figure 2 was used to develop the proposed Safety Information Management System. The methodology is based on the *Action Research* technique. Action research is an iterative research method in which the researcher investigates the problem domain, identifies the problem, gets involved in introducing some changes to improve the situation and evaluates the effects of those

changes. Action research combines theory and practice and is very useful in solving a practical problem in a real setting (Azhar, 2007). The following steps will briefly explain the methodology and system development efforts.

In the first phase, the existing safety information management systems typically used in the construction firms were critically evaluated to identify their capabilities and limitations. This was done via a questionnaire survey targeted at general contractors, specialty contractors and subcontractors. The sample size was 550 firms and the response rate was 31%. The results revealed that most of the surveyed firms do not effectively utilize safety information for planning and decision-making due to its non-availability in useful forms and formats to the decision makers. The results of the questionnaire survey also served as a guide map to capture user requirements for developing the proposed safety Information Management System (Phase 2).
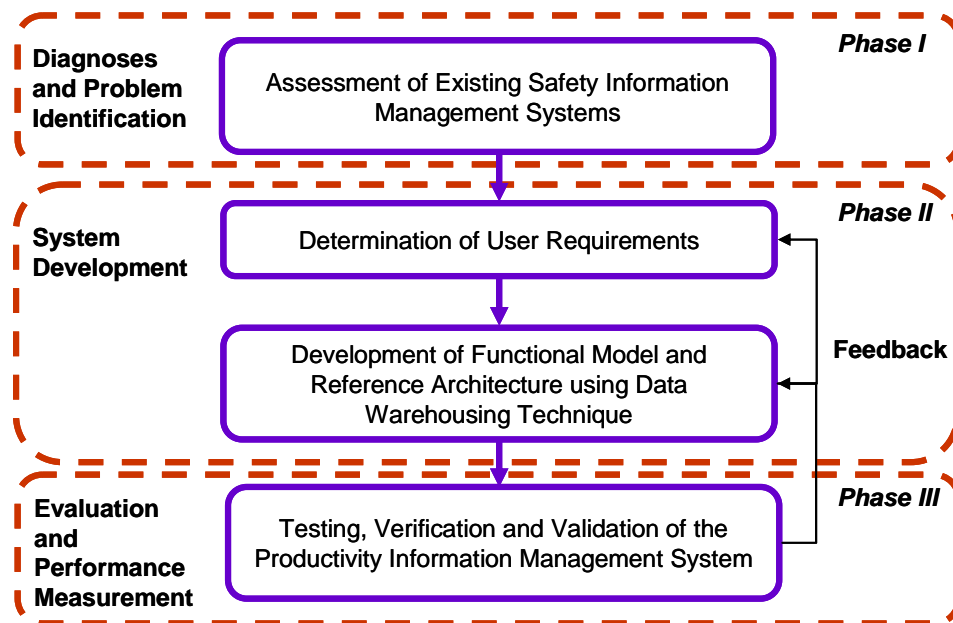

Figure 2: Research Framework

Based on the user requirements, a functional (or logical) model of the proposed system and its corresponding reference architecture were developed using the data warehousing technique. The functional model represents how the system would work while the reference architecture dictates its physical and technical implementation using different software and hardware tools. Figure 3 illustrates the data warehousing development framework used for *SafeMan*.

There are two approaches to build a data warehouse. In the first approach, stand-alone data marts (a subset of a data warehouse categorized according to a functional area (such as Safety Information or Cost Control) are developed and later integrated in to an enterprise-wide data warehouse. In the second approach, a complete warehouse in the form of distributed data marts is build. These data marts are populated with data either at the time of initial development or at different stages depending on the availability of time and resources (Ahmad and Azhar, 2005). The first approach is adopted in this research as it fits with the scope of this research, and is relatively simple. The data mart schema design constitutes the major portion of the *SafeMan* system development and hence will be explained in detail. The basic information about other design modules (such as data extraction schemas, OLAP analysis tools, etc.) can be found in the following reference (Azhar, 2005).
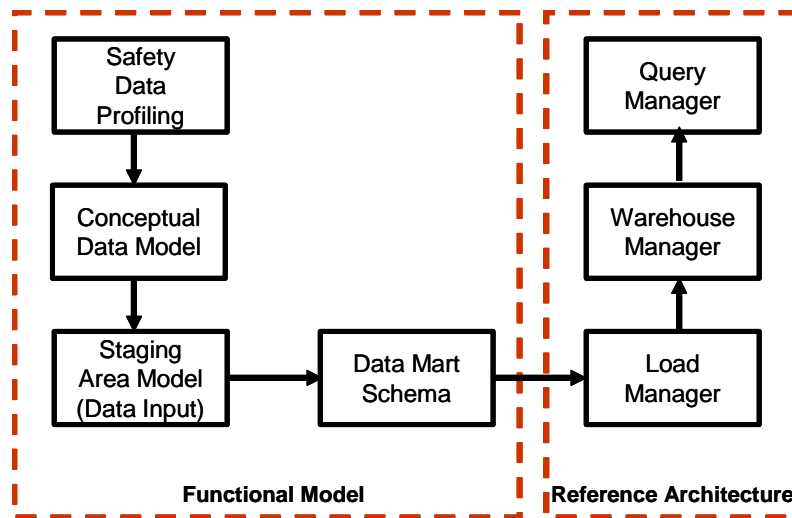
Figure 3: Data Warehouse Development Framework

The data mart (or data warehouse) schema design essentially consists of three steps as follows: (1) identifying Facts and Dimensions; (2) designing Fact and Dimension tables; and (3) designing data mart schemas.

Facts represent quantitative data about a business entity/transaction (e.g. Injuries count, Accidents claims amount, etc.) while Dimensions contain descriptive data that reflect the dimensions of that entity (e.g. time, project size, location, etc.). Once the Facts and Dimensions are established they can be organized in the form of Fact and Dimension tables. The Dimension tables are connected with the Fact table via foreign keys thus forming a data mart schema. There are three main types of data mart design schemas: the *Star schema*, the S*nowflake schema* and a hybrid of the two, the *Starflake schema* (Ahmad *et al.*, 2004). The Star schema as shown in Figure 4 is adopted in this research due to its simplicity. The data stored in the Star schema must be denormalized (a technique to move from higher to lower normal forms of database modeling in order to speed up the database access and thus query processing). Due to demormalized data, the Star schema is considered as very efficient in answering complex queries (Ahmad *et al.*, 2004).

A prototype Safety Information Management System named *SafeMan* was developed to verify the functional model and to demonstrate its capabilities to the users. The reference architecture of *SafeMan* is shown in Figure 5. The load manager and warehouse manager were developed using *MS Access™*. The load manager extracts operational data stored in spreadsheets and applications databases (e.g. payroll database, claims database, etc.). These data are validated (a process which ensures that data values are correct and within the permissible range, e.g. a valid range of number of hours in a day is from 0 to 24, thus any value below 0 or above 24 would be considered as invalid), filtered (to remove redundant data) and integrated (a process of combining data from two or more than two data sources) before final storage. The query manager was designed using *Power OLAP™*. The users can retrieve stored data via query manager to perform different OLAP operations (e.g., n-Dimensional analysis, Slice and dice analysis, Drill down analysis, etc). The *Power OLAP®* exports selected data to *Xcelsius™* for preparation of interactive charts and graphs.

For verification and validation purposes, the prototype *SafeMan* system was implemented in a construction organization in Miami, Florida. Data of six selected projects were fed in the prototype system and its performance was evaluated by a Focus Group consisted of senior management personnel of the selected organization.
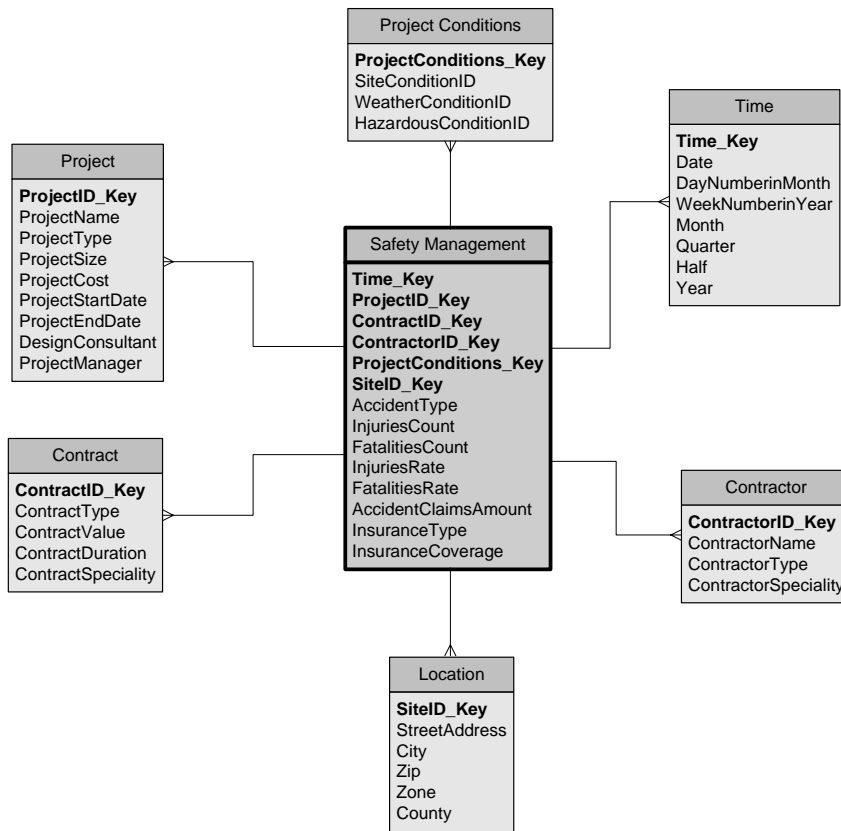
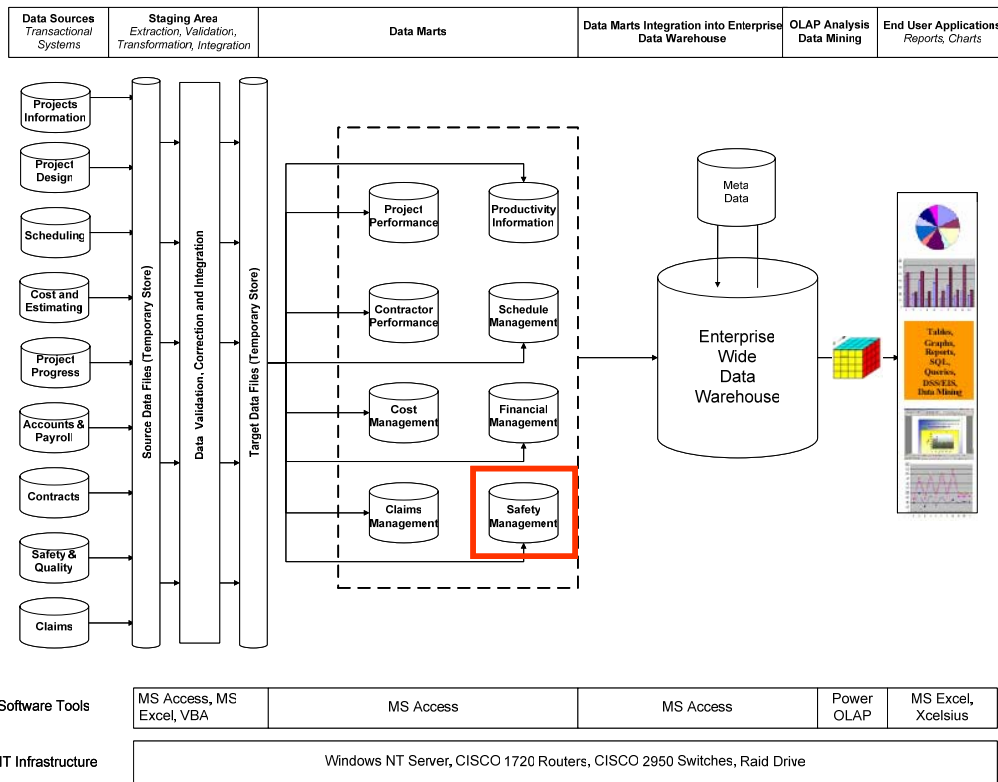Figure 4: Star Schema used for Safety Information Management Data mart

Figure 5: Reference Architecture of the *SafeMan* System

4. System Demonstration and Evaluation

The *SafeMan* extracts input data from existing spreadsheets and databases as shown in Figure 5. The input data are validated, filtered and integrated before storage in the safety management data mart. The safety data stored in the data mart are then analyzed using regression and other statistical analysis models. These models are used to calculate different safety related parameters as listed in the Fact table of Figure 5. More details about these models will be published in a separate paper. The users can analyze safety data with respect to various projects dimensions such as project size, type, location, site conditions, etc. or can perform scenario analysis to analyze site safety under different projects conditions.
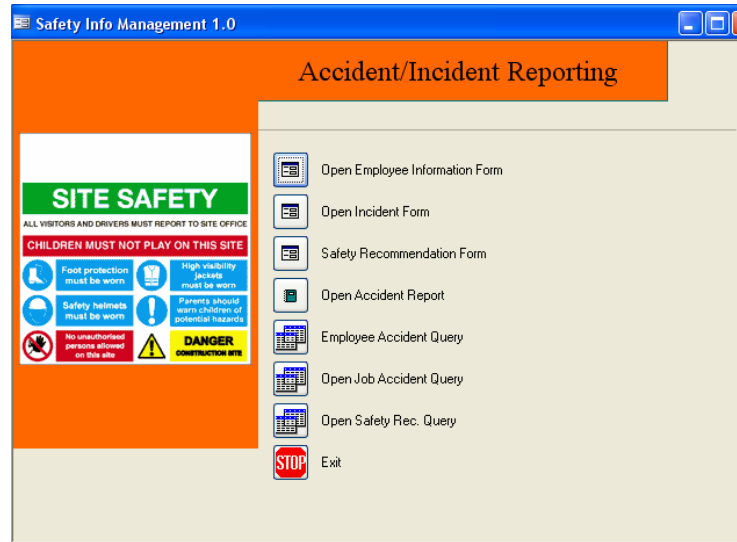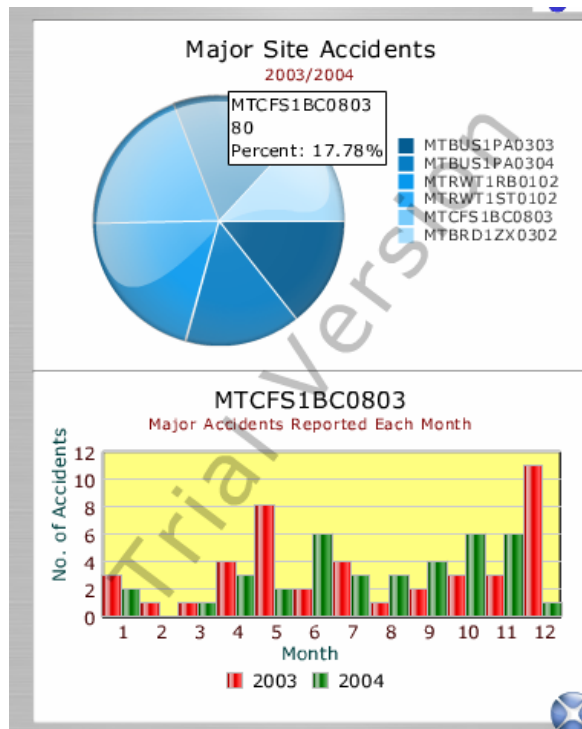

Figure 6: A Sub-menu of the *SafeMan* System

Figure 7: Slicing Site Accidents Data for Project "MTCFS1BC9803" and Dicing for Years 2003 and 2004

Figure 7 illustrates the Slice-n-Dice analysis of safety records of a selected project. This type of analysis enables user to view both summarized and the detailed data to explore accidents trend in a particular trade or during a particular month. Besides this example, the *SafeMan* can perform dimensional analysis to analyze safety performance of different subcontractors engaged in different trades in different projects. The system is still under development and more functionalities will be added in the near future.

The opinions of the Focus group before and after testing this system are illustrated in Table 1. The results indicate that the significant error (p) in all cases is below 0.05 ($<5\%$). This means that the results are statistically significant, i.e. significant improvements were recorded after the implementation of the *SafeMan* system.

Table 1: Comparison in Performance Before and After using the *SafeMan* System

Comparison Criteria	Mean of Responses		Sum of Squares (X^2)	Significant error (p)
	Without *SafeMan*	With *SafeMan*		
Ease in data access	2.57	4.53	7.96	0.02
Data quality	2.12	4.76	7.12	0.03
Support for every day decisions	2.59	4.89	8.96	0.00
Support for short and long term planning	1.60	4.56	10.16	0.01

5. Concluding Remarks

A Safety Information Management System (*SafeMan)* is developed using the data warehousing technique. The *SafeMan* can be used for performing (1) multidimensional data analysis; (2) slice-n-dice analysis; and (3) scenario analysis. These analyses can help decision-makers to evaluate safety performance and project planning and execution mistakes leading to site accidents. Based on these analyses, corrective actions can be suggested and insurance claims can be more easily resolved. The main objective of this paper was to examine the suitability of data warehousing technique for safety information management which was successfully achieved. The future papers will illustrate the regression and multivariate models used in the *SafeMan* system for safety information analysis.

References
Ahmad, I., and Azhar, S. (2005). "Implementing Data Warehousing in the Construction Industry: Opportunities and Challenges". *Proceedings of the Third International Conference on Construction in the 21st Century (CITC-III)*, Athens, Greece, September 15-17, pp. 863-871.
Ahmad, I., Azhar, S., and Lukauskis, P. (2004). "Development of a Decision Support System using Data Warehousing to Assist Builders/Developers in Site Selection." *Automation in Construction,* 13(4), 525-542.
Ahmed, S.M.; Azhar, S.; and Forbes, L.H. (2006). "Costs of Injuries/Illnesses and Fatalities in Construction and their Impact." Proceedings of the CIB W99 International Conference on Global Unity for Safety and Health in Construction, Beijing, China, June 25-27, 363-371.
Azhar, S. (2005). *Information Systems to Support Decision-Making in Construction Owner Organizations: A Data Warehousing Approach*, Ph.D. Dissertation, Florida International University, Miami, Florida, USA.
Azhar, S. and Ahmad, I. (2006). "A Case Study of Data Warehousing Implementation in a Public Construction Owner Organization." Construction Data Warehouse, 2nd ed, Edited by Sang-Hoon Song, Seoul National University, Seoul, Korea, 15-31.
Azhar, S. (2007). "Improving Collaboration between Researchers and Practitioners in Construction Research Projects using Action Research Technique". *Proceedings of the 43rd ASC National Annual Conference* (on CD-ROM), Flagstaff, Arizona, April 12-14, 2007.
Bureau of Labor Statistics. (2007) "Fatal and Non-Fatal Occupational Statistic from 1992-2005". Online at http://www.bls.gov. Accessed on April 01, 2007.
Mincks, W.R. and Johnston, H. (2004). *Construction Jobsite Management*, 2 ed., Thomson Delmar Leaning, NY.