Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (21): 4560-4564, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.4560.4564

Study on Application of Data Warehouse in Medical Insurance Fund Mis

Zhou Huiwen
School of Public Finance and Administration,
Zhejiang University of Finance and Economics, Hangzhou, 310018, China

Abstract: There are relatively few institutions that have developed data warehouses in medical insurance fund management information system. This study provides an overview of the application of data warehouse in the system. Data warehouse in the system could play a key role in conducting data integration, information analysis, data mining and quantitative forecasting. The proposed management information system with data warehouse could automatically assign appropriate inspection and renewal evaluations for each insurance units and estimates associated integrated payments, resulting in effective and practical medical insurance fund management from various perspectives, with corresponding levels of detail.

Key words: Data warehouse, MIS, medical insurance, fund management

INTRODUCTION

The data that medical insurance fund management need is always related with the multi-dimension computation and statistics. The multi-dimensional analysis based on the multi-dimensional data is the primary coverage of the analysis. On the other hand, the Data Warehouse (DW) technology which is used to extract and analyze DSS data, create the method of multi-dimensional data analysis and processing. However, in order to support the complex management analysis, how data warehouse system conducts the effective organization and management to mass data mainly relies on the design of Logical Data Model (LDM) in the data warehouse system (Wang, 2010; Seabo *et al*., 2011). The DW technology's multi-dimensional characteristics satisfy the DSS requirements of data analysis. At the same time, DW overcomes disadvantages of the bad data organization and the low utilization rate. Now, the priority is to extracting the useful information from the mass data of Medical Insurance Fund Management Information System (MIFMIS) and applying it into the fund management. Therefore, the design of MIFMIS based on the data warehouse is extremely valuable.

DESIGN AND ESTABLISHMENT

System design principles:

- **Principle of universalization:** There are differences among local medical insurance fund management branches in organizational structures, business segments and focuses, RDBMS that running on OLTP and the types and formats of the data supporting insurance fund management analysis and decision system. These differences are considered within the general design of the medical insurance fund MIS
- **Principle of expandability:** Along with the change of service, the range of information of medical insurance management system will change. The expandability of system flexible functions should be considered in the well designed systemic structure of information service platform. This means while maintaining system architecture and the diagnosis logic of original services, the system can satisfy theme analysis and functionality expansion
- **Principle of compatibility:** Using different data collection and conversion methods, the medical insurance information system based on data warehouse needs to firstly transform the historical data stored in various types of traditional database and business systems into the medical insurance fund management data warehouse after consistency check and clean up (Huang and Hu, 2010)
- **Principle of standardization:** The development of the MIFMIS should meet the requirements of the relevant regulations of the fund management authorities. All the standards such as index system standards, data interface standards, network communication standards, business norms, information data entry, information classification and coding standards and relevant technical standards have to be uniformed and follow the fund management regulations

Establishment process of MIFMIS: In the establishment of MIFMIS, the OLAP technology first aggregates data and then separates the data into dimensions and measures. Based on the dimensions and metrics analysis, the results present to users in an intuitive form.

Data warehouse stores and integrates the management theme data. OLAP analyzes multi-dimensional data. Data Mining (DM) is used to mine knowledge from data warehouse. The model library can be combined with multiple generalized models for the management and monitoring (Prat *et al*., 2011). The MIFMIS is able to use knowledge reasoning approach for qualitative analysis and finally make a comprehensive monitoring analysis.

There are distinctions and connections among data warehouse, data analysis and data mining while they permeate and supplement each other. Therefore, by integrating these technologies, MIFMIS could conduct more efficient monitoring and management.

DATA WAREHOUSE

System architecture of data warehouse: The data warehouse is a type of organization form of the stored data. From the perspective of the logical structure, the data can be divided into 3-4 different layers. Initial raw data is from traditional databases. Firstly, according to the requirements of decision-making support themes, the raw data forms the current basic data layer. After that the data forms the comprehensive data layer (the mild comprehensive data layer and the highly comprehensive layer), according to integrated decision-making requirements. This current basic data layer transfers into the historical data layer by the time control mechanism as the time passes. These data layers are all organized by metadata which is the core of the data warehouse. From the perspective of physical structures, the organization of data warehouse can be divided into the multi-dimensional organization (spatial hyper-cube), the database related forms of organization and the virtual storage. The data warehouse system is used to extract, transform, filter, clean and process the raw data into the data warehouse. Servers update the restored data in the data warehouse and use related software to integrate the data (El-Sappagh *et al*., 2011; Nielsen, 2011). The DWS is usually composed by data warehouse, warehouse management and analysis tools (Fig. 1).

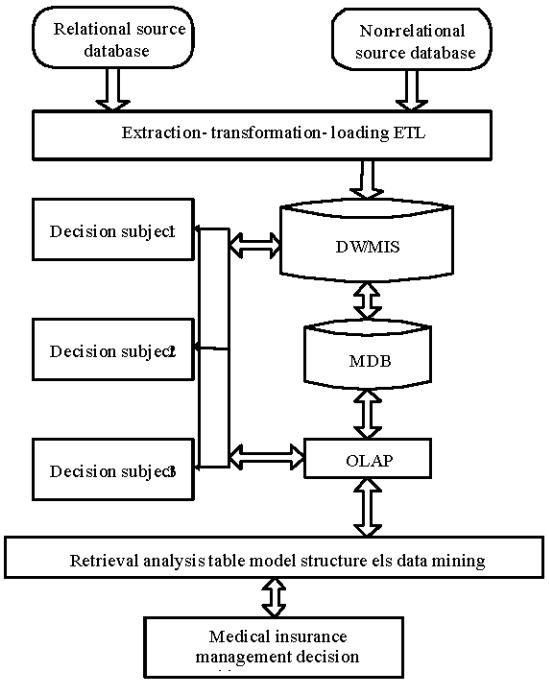

Fig. 1: System architecture

- **The data source:** In Fig. 1, the most left side is the data source which includes database data source and non-database data source. The database is generally heterogeneous and interconnected through a network. The data source is the basis of the data warehouse system. Data comes from different data sources such as hospital data, insurance policy holders' data and medical insurance institutions' data must be extracted and transformed to establish data warehouse
- **integration of the data sources:** Integration process of multiple data sources includes extraction, transformation and loading process. Data integration is an important section of building a medical insurance fund management data warehouse. Users extract data from the data source, though conversion and cleaning, then load the data in to data warehouse according to pre-defined models of the data warehouse
- **Central data warehouse:** The current data warehouse core is still a database system under the management of RDBMS. Due to the mass of data and query complexity, there is still a problem to determine whether the system performance is good or not. Therefore, in order to improve the performance, there are the three measures in the data warehouse: (a) Use parallel system structure, (b) Specially deal with the frequent use data, (c) Define the query results as a view and calculate views in advance to store in the data warehouse which can be read out when needed without temporary calculated. In the above three

measures, parallel processing is a common technology in database, supporting by all major RDBMS products; latter two measures are developed with the rise of the data warehouse while major RDBMS products have increased the relative functions

- **Data mark:** In order to solve the contradiction between flexibility and performance, the data warehouse adds related data marks in the architecture. Data mark stores pre-calculated data for a particular user to meet the performance needs of users. Data mark is a small data warehouse for specific users or specific themes. The system uses information mining technology, integrates client environments, filters out the information that users are most interested in to meet users information needs by screening information, induction and summarizing. Then, the system automatically generates users' preference data and saves to the related data marks for subsequent data analysis
- **On-line analytical processing (OLAP) server:** The OLAP server locating between the clients and the data warehouse is responsible for the management of the data model and its data. The server is actually a powerful multi-dimensional analysis engine. It extracts the data from the data warehouse and effectively integrates the analytical data based on organized multi-dimensional model in order to do multi-angle and multi-level analysis and explore the trend
- **Front-end analysis tools:** The front-end analysis tools include a variety of query tools, data analysis tools, reporting tools, data mining tools and a variety of applications development tools based on the data warehouse and data marks. The data analysis tools mainly focus on OLAP server. Reporting tools and data mining tools focus on data warehouse. All the above front-end analysis tools run on the client terminal and the main function is to provide a multi-dimensional data query and analysis in order to achieve the purpose of decision-making support

Medical insurance data design: For the purposes of social security data mining, the data is classified into several dimensions to presenting basic categories of data and analysis levels. Then, each dimension is cut into various information granularities to presenting the characteristics of the current category. This division cannot have too much excessive levels or too small affairs. Otherwise it is not only difficult to design the mining algorithm, the final results also lack of the overall representativeness and significance.

- **Annual dimensions:** Concerning the starting year of the insurance, the grain size is divided from the medical insurance system starts to the current year that conducted data mining calculation. However, due to the policy holders' status or insurance type change by special reasons, the information granularity can be determined by a number of stages, until the death of policy holders
- **Date dimension:** The granular can be divided by the 12 months, focusing on different months' data changes to mine the laws and rules
- **Gender dimension:** The male and female can be divided into two information granularities. It is meaningful to distinguish gender in order to determine the different characteristics of the medical insurance
- **Wage dimension:** From 0 to 10,000 RMB, we treat 500 RMB as a partition of order. With the wage fluctuations, the amount of the related income will be adjusted. Furthermore, there will be a relation between the insured amount and the wage level
- **Disease dimension:** Classified monitoring can be conducted according to different disease types, such as common diseases and major diseases. For those holders with specific diseases, the fixed medical institutions, the required special drugs and billing need to be checked, then payments could be completed
- **Age dimension:** The best way to classify ages is according to insured person's data of birth, because age will gradually increase which needs higher refresh requirements for the database and more complicated calculations. The number of year can be defined as the retirement time scale which will help the system to determine the type and start time of premium payments
- **Unit dimension:** The unit can be divided into enterprises, institutions, organizations, associations, individuals and streets. For different units, medical insurance investment limits and types will have differences which can impact of the effects of data mining
- **Personnel dimension:** The personnel dimension can be divided into practitioners, retired and unemployed. In accordance with the personnel natural conditions, the classified mining can provide an effective basis for the insurance policy to different groups of people in the medical insurance
- **Payment dimension:** Dimensions also can be further divided into sub dimensions based on the difference between insurances. Medical insurance payment information, especially, can be divided into treatment,

hospitalization and purchase of medicines along with the record of time, etiology, treatment locations and bills. This is helpful to track the history of policy holders, timely analysis of health status and prevent unauthorized access of the medical insurance by others (Huang, 2012)

APPLICATION

Data model analysis: Due to the keeping changing of comprehensive analysis of the medical insurance fund, data storage in the warehouse must use the theme domain methods and small business units to organize in order to meet the data warehouse flexibility and needs changes (Ehmke *et al*., 2011). Meanwhile, all information systems have the characteristics of integrity, constitutive property, hierarchical, relativity and variability. These characteristics of the design reflect the goal of the logical structure of the data warehouse system and also correctly reflect the objective requirements of the target systems.

Through the analysis of business needs, the business of the medical insurance fund can be divided into different categories by five theme domains. They are insurance policy holders, the insurance units, integrated payments, medical treatment and overall revenue and expenses of the fund. Insurance policy holders theme contains basic information associated with the insured population, such as the natural and social properties of the insured population (name, age, occupation, etc.). The insurance unit theme contains basic information related to the insurance units, such as natural and social properties of the insurance units (unit type, economy type, industry classification, etc.). The integrated payment status analysis includes insurance unit and individual contributions. The fund's revenue and expenditure analysis includes the insurance fund's payments and balance analysis of every type of insurance (Ghannouchi *et al*., 2010).

The data model of medical insurance fund information system is a multi-dimensional data model, including the star model and snow-flake model. The star model contains a theme fact table and several non-standardized described dimension tables with the outside connections (Nielsen, 2011). The medical insurance fund management data maintains in the fact table and the dimension data maintains in the dimension table. Each dimension table directly connects with a fact table through a key word. The dimensions are the classified information such as time, the location, the hospital, sickness type and so on. The snowflake model is one kind of expansion form of the star model. Dimension tables are broken down into the primary dimension table that directly related with the fact table and the associated dimension table that related with the fact table and primary dimension tables. Comparing with the star model, snow-flake model is easier to explore data dimensions, but the processes by computer are more complex (Schuff *et al*., 2011).

Thematic data mining: The medical insurance fund management accumulated a large amount of business data in the process of informatization construction. The data size of big cities have more than Terabytes (TB) level; At the same time, the difficulty and process complexity of medical insurance fund management reform determine that the management authority needs the extraction and analysis of history insurance data into useful information. Data mining is typically used in a massive data set to find the indirect, hidden and new rules or laws in order to assistant decision-making.

Data mining technology has the advantage of exploring the useful information from the mass medical insurance data by limited steps of collation, analysis, inference, comparison and other analytical methods. Data mining is also a highly inductive information technology. It reorganizes the mass data set into the operational conclusions and regulations which can be used to guide the management decisions of the medical insurance fund. Common algorithms of data mining include: association rules, cluster detection, decision trees, support vector machines, neural networks and genetic algorithms. The data mining tools like SAS and IM8 support those algorithms including decision tree, cluster analysis, regression analysis, neural networks and so on.

The data mining thematic analysis of medical insurance fund management supports the monitoring and analysis of the medical insurance business data which includes that of medical insurance policy holders, medical insurance designated hospitals and medical insurance fund revenue and expenditure.

CONCLUSION

Data warehouse is the central information repository required by medical insurance decision maker. With the help of the multi-dimensional schema, the data warehouse allows the user to explore the measures from various perspectives of analysis. This management information system uses the data warehouse technology to successfully solve data integration, information analysis, data mining and data prediction problems in the heterogeneous business systems of medical insurance fund management. According to the themes of decision analysis, those insurance fund management data is

reorganized through the data warehouse. Then, the managers can conduct decision analysis based on different concerned themes and topics. The application of the data warehouse improves the work efficiency of medical insurance fund managers through online dynamic query analysis. This system optimizes and improves the existing workflow, monitoring and analytical methods in medical insurance fund management analysis.

REFERENCES

Ehmke, J.F., D. Grobhans, D.C. Mattfeld and L.D. Smith, 2011. Interactive analysis of discrete-event logistics systems with support of a data warehouse. Comput. Ind., 62: 578-586.

El-Sappagh, S.H.A., A.M.A. Hendawi and A.H. El Bastawissy, 2011. A proposed model for data warehouse ETL processes. J. King Saud Univ. Comput. Inform. Sci., 23: 91-1040.

Ghannouchi, S.A., K. Mabrouk and S. Ghannouchi, 2010. Proposal of data warehouse in the context of healthcare process reengineering. Bus. Process Manage. J., 16: 688-712.

Huang, F., 2012. Study on data mining in social insurance information system. Sci. Technol. Innov. Herald, 30: 53-54.

Huang, J.H. and B. Hu, 2010. Study on applying delaware technique in social insurance. China Comput. Commun., 3: 124-125.

Nielsen, A.C., 2011. Data warehouse for assessing animal health, welfare, risk management and communication. Acta Vet Scand., Vol. 53 10.1186/1751-0147-53-S1-S3

Prat, N., I. Comyn-Wattiau and J. Akoka, 2011. Combining objects with rules to represent aggregation knowledge in data warehouse and OLAP systems. Data Knowl. Eng., 70: 732-752.

Saebo, J.I., E. K. Kossi, O.H. Titlestad, R.R. Tohouri and J. Braa, 2011. Comparing strategies to integrate health information systems following a data warehouse approach in four countries. Inform. Technol. Deve., 17: 42-60.

Schuff, D., K. Corral and O. Turetken, 2011. Comparing the understandability of alternative data warehouse schemas: An empirical study. Decision Support Syst., 52: 9-20.

Wang, T., 2010. Data model design of hospital data warehouse. Comput. Technol. Dev., 20: 191-195.