DECISION SUPPORT SYSTEM FOR DEPARTMENT OF HEALTH BY USING DATA WAREHOUSE & OLAP TECHNIQUES

Than Lin Tun, Nan Saig Moon Kham
University of Computer Studies, Yangon, Myanmar
thanlinntun@gmail.com, moonkhamucsy@gmail.com

Abstract

Since the introduction of computer system, decision support has evolved greatly. Originally computer systems produced paper reports. The emerging technologies of data warehousing and OLAP have changed the way that organization utilize their data. Data Warehousing and OLAP have created a new framework for organizing corporate data, delivering it to business end users, and providing algorithm for more powerful data analysis. The system is analyzed morbidity data to load into a data warehouse. The data are transformed into a data cube in dimension of Time, Disease, Location and Patient. Then, the required data are analyzed by multidimensional OLAP model to provide necessary analysis result to the decision maker.

1. Introduction

The Department of Health, one of the seven departments under the Ministry of Health is responsible for providing health care services to the entire population in the country. Under the supervision of the director general and deputy director general , there are 9 directors who are leading and managing the following divisions. 1. Administration, 2. Planning, 3. Public Health, 4. Medical care, 5. Disease control, 6. Food and Drug Administration, 7. National Health Laboratory, 8. Occupational Health and 9. Nursing. Function of disease control division cover prevention and control of infectious diseases, disease surveillance, outbreak investigation and response and capacity building.[11]

To fulfill the need of integrated national health information system ensuring timely, reliable and accurate information based on minimal essential data set, the Health Management Information System (HMIS) was established 10 1995. The main objective are to ensure minimum essential information of prioritized health projects are integrated in the national health information system, to generate and report health information in the course of implementation of the National Health Plans for timely and effective monitoring and evaluation and to reduce the data collection burden for basis health staff. HMIS includes community based as well as institutional based information as a means to support making evidence based decisions in policy design, planning and management so as to improve overall health system performance. Hospital reporting is another facet of health information service well established through monthly collection of hospital morbidity and administrative information from public hospitals. Morbidity information which is individual case summaries with analysis of all discharges and deaths is processed at the central office (Department of Health Planning). [12]

In order to implement the morbidity, it is essential to know the information of morbidity in respect with the time, disease, patient and location . This system has been approached by using the data warehouse and OLAP technology to maintain and update the required information for decision making.

The rest of the paper is organized as follows: Section 2 describes related work. Section 3 presents Background Theory of Data Warehouse & OLAP. Section 4 describes Implementation of the system & Results. Section 5 gives Conclusion of the system.

2. Related Work

Innmon [7,8] has been presented a data warehouse is a repository of integrated information, available for querying and analysis. As relevant information becomes available or is modified, the information is extracted from is source, translated into a common model e.g., the relation model), and integrated with existing with the data warehouse. Jarke [6] presented a metadata modeling approach which enables the capturing of the static parts of the architecture of a data warehouse.

Marotta [2] presented a set of schema transformation primitives for relational DW design. Widom [9] presented a general research problem in the data warehouse. The database community is devoting increasing attention to the research themes concerning data warehouses (DWs); nevertheless, the crucial issues related to DW design have not been deeply investigated yet.

3. Data Warehouse and OLAP

Data warehouses generalize and consolidate data in multidimensional space. The construction of data warehouse involves data cleaning, and data transformation and can be viewed as an important preprocessing step for data mining. Moreover, data warehouse provide on- line-analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities. Data warehousing provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions. A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision making process. [5]

OLAP stands for "On-Line Analytical Processing" and describes a class of technologies that are design for ad-hoc data access and analysis.[13]

OLAP uses a multidimensional view of aggregate data to provide quick access to strategic information for further analysis. OLAP enables analysts, managers, and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information. OLAP transforms raw data so that it reflects the real dimensionality of the enterprise as understood by the user. OLAP enables decision-making about future actions. A typical OLAP calculation is more complex than simply summing data. OLAP transforms Data Warehouse data into strategic information [1]. OLAP includes a set of operations for manipulation of the dimensional data organized in multiple levels of abstraction. As decision-makers exercise more advanced OLAP capabilities, they move from data access to information to knowledge [5].

3.1 Warehouse Architecture

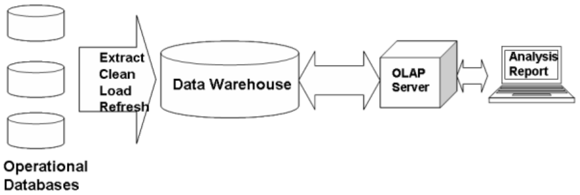

Figure.1. Data Warehousing Architecture

Figure 1 shows a typical data warehousing architecture. It includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system [3].

Data warehousing systems use a variety of data extraction and cleaning tools, and load and refresh utilities for population warehouse.

After extracting, cleaning and transforming, data must be loaded into the warehouse. Typically, batch load utilities are used for this purpose. In addition to populating the warehouse, a load utility must allow the system administrator to monitor status, to cancel, suspend and resume a load, and to restart after failure with no loss of data integrity. The load utilities for data warehouses have to deal with much lager data volumes than for operational databases.

Refreshing a warehouse consists in propagating updates on source data to correspondingly update the base data and derived data stored in the warehouse. Usually, the warehouse is refreshed periodically (e.g., daily or weekly). Refresh techniques may also depend on the characteristics of the source and capabilities of the database servers, extracting an entire source file or database is usually too expensive, but may be the only choice for legacy data sources.

3.2 Multidimensional Model

Data warehouses and OLAP tools are based on a multidimensional data model.The multidimensional data model is based on the key concept cube, dimension and hierarchy. The cube is the concept that describes the whole of the data which are presented along labelled edges of that cube, the dimensions. Whereas hierarchies define the way in which dimensions are grouped. [5]

3.3 Data Cubes

A data cube allows data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts. It is a multi-dimensional data structure, a group of data cells arranged by the dimensions of the data. For example, a spreadsheet exemplifies a two-dimensional array with the data cells arranged in rows and columns, each being a dimension. A three-dimensional array can be visualized as a cube with each dimension forming a side of the cube, including any slice parallel with that side. Higher dimensional arrays have no physical metaphor, but they organize the data in the way users think of their enterprise. Typical enterprise dimensions are time, measures, products, geographical location, sales channels, etc. It is not rare to see more than 20 dimensions. However, the higher the dimensions the more complex the manipulation and data mining on the cube become, and the more space the data cube may become.[10]

3.3.1 The Lattice for Morbidity Data Cube

The lattice of cuboids is referred to as a data cube. In this system, a lattice of cuboids forming a data cube for the dimensions time, disease, location and patient. It can be computed for this data cube. It has three attributes, time, disease and location as three dimensions and morbidity_ in _units as the measure, the total number of cuboids, or group-by's that can be computed for this data cube is $2^3 = 8$. The possible group-by's are the following: {(time, disease,location),(time,disease),(time,location),(disease,location),(time),(disease),(location),( )}, where ( ) means that the group-by is empty. It contains the total sum of all productions. This group-by's form a lattice of cuboids for the data cube. The base cuboids contains all four dimensions.

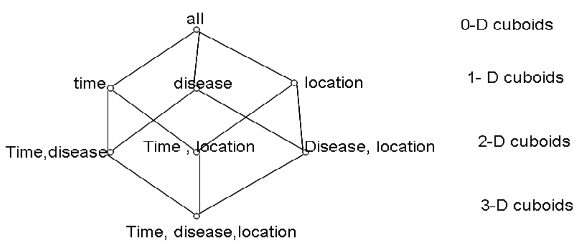

Figure. 2. Lattice of Cuboids for DOH Data

3.3.2 OLAP Operations

(i) Roll-up

Some dimensions have a hierarchy defined on them Aggregations can be done at different levels of hierarchy. Going up the hierarchy to higher levels of generalization is known as roll-up. For example, aggregating the dimension up the hierarchy (day - month - quarter) is a roll-up operation.

(ii) Drill-down

These operation traverses the hierarchy from lower to higher levels of detail. Drill-down displays detail information for each aggregated point.

(iii) Slicing-dicing

This operation involves selecting some subset of the cube. For a fixed attribute value in a given dimension, it reports all the values for all the other dimensions. It can be visualized as a slice of the data in a 3D-cube.

3.4 Warehouse Schema

Most data warehouse use a star schema to represent the multidimensional data model. The database consists of a single fact table and a single table for each dimension. Each tuple in the fact table consists of a pointer (foreign key-often uses a generated key for efficiency) to each of the dimensions that provide its multidimensional coordinates. Each dimension table consists of columns that correspond to attributes of the dimension. Figure 3 shows a star schema of the system.

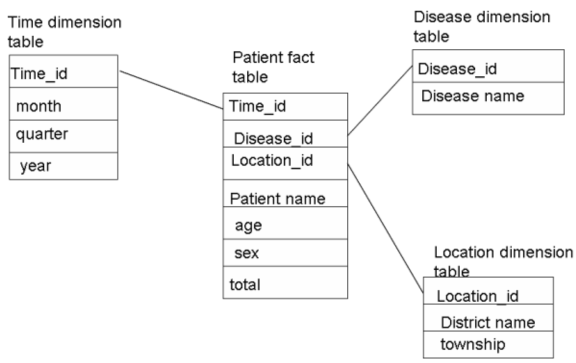

Figure3.Star schema for DOH data warehouse

3.5 Overview of the System Design

Overview of the system is described in the following Figure 4. Firstly, data from operational data sources are loaded into the data warehouse. To load data into data warehouse, the system first opens the operational database, extracts into star schema then loaded into the data warehouse. After loading warehousing, data are analyzed with the OLAP operations for the user requirement of the query report. After reporting, the query, user can request new query for another analysis report or exit the system.

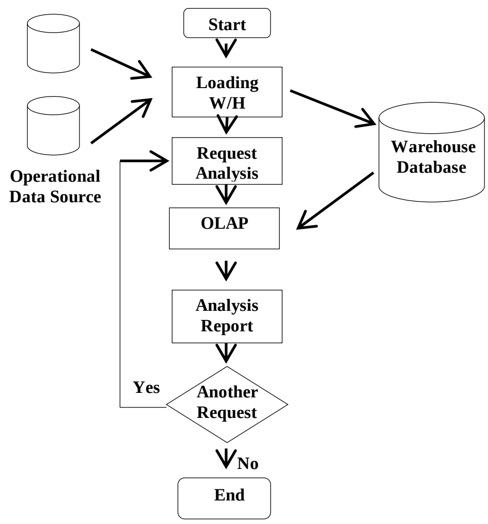

Figure. 4. Overview of the System

4. Implementation of the System & Result

In this system, the DSS model is implemented by integrating data from the different data sources. This system is applied same operational database. Operational database can also come from many different physical sources. In this paper, operational database is obtained from Department of Health. In the current study, data recorded from Department of Health of historical data was used. It has been recorded from yearly 2004 to 2008. This system is applied only loading process, ad-hoc query process, refresh process and analytical reporting for decision making. OLAP may need to access different analytical processing. It may need to access different cuboids for different queries.

According to the lattice of cuboids, the system can process eight query reports for the department of health. They are the following reports:-

(1) Analysis report by – time, disease, location
(2) Analysis report by - time, disease
(3) Analysis report by – time, location
(4) Analysis report by – disease, location
(5) Analysis report by – time
(6) Analysis report by – disease
(7) Analysis report by – location
(8) Analysis report by – all

Example query is querying for patients amount of Tuberculosis for East district on yearly 2004, 2005, 2006, 2007 and 2008.
The table 1 is Sample Result of above query form for Department of Health.

Disease Name	District Name	Township	Year	Total Patient
Tuberculosis	East	Dawbon	2004	126
Tuberculosis	East	Tharkata	2005	210
Tuberculosis	East	Tarmwe	2006	225
Tuberculosis	East	Yankin	2007	186
Tuberculosis	East	Botahtaung	2008	113

Table 1. Result of Normal query for Department of Health

As roll up operation for the above query with the time dimension to value 2005, the table 2 is sample result of that roll up operation for department of health.

Disease name	District name	Year	Total patient
Tuberculosis	East	2005	758
Tuberculosis	West	2005	668
Tuberculosis	South	2005	587
Tuberculosis	North	2005	995

Table 2. Result of roll up for department of health

AS the drill down operation for the query resulting table 1 with lowering hierarchy of location dimension and time dimension, the table 3 is sample result of that drill down Operation for department of health.

Disease name	District name	Township	Year	Quarter	Month	Total patient
Tuberculosis	East	Dawbon	2004	1	March	13
Tuberculosis	West	Latha	2004	1	January	10
Tuberculosis	North	Taikkyi	2004	2	April	15
Tuberculosis	South	Dala	2004	3	August	16

Table .3. Result of drill down for Department of Health

The results of the system are as shown in the following figures. The user interface of our system includes initial loading process and querying facilities as well as display of query result for decision making.

Figure 5 represents the analysis report of time, disease and location by Total patient of system.

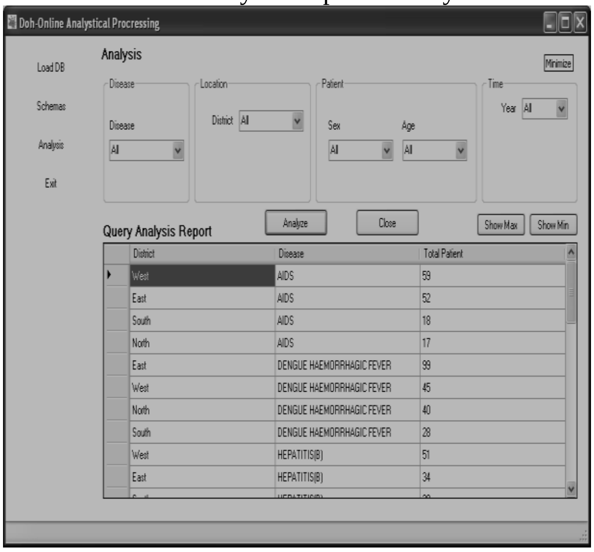

Figure 5. Analysis report by All

As a facility of the system, maximum caused disease within user's query can be calculated for decision making. Figure 6 shown the result of disease which cause maximum patient amount among all districts, altogether with the patient amount caused by the resulting disease. User can query the maximum caused disease for any query of any dimension filter.

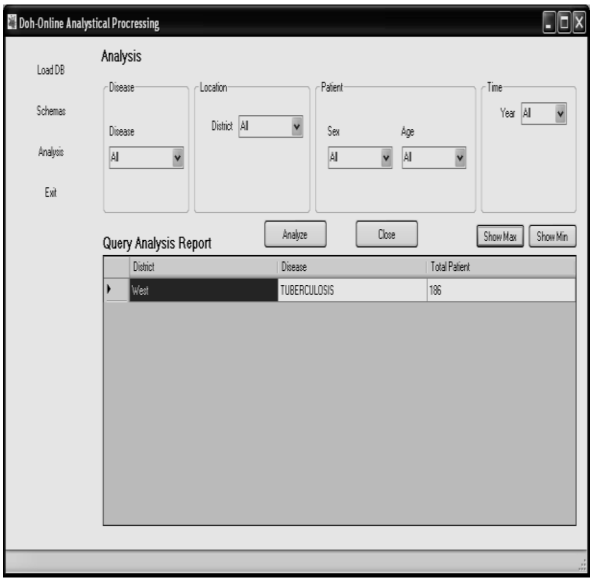

Figure 6. Report of maximum causes disease

5. Conclusion

This system is built for users who are in medical admin officers of department of health. Information in a data warehouse is organized in star schemas, which simplify user understanding of the data, maximize performance for decision support applications, and require less storage for a large database. Our approach can applied to systematic organizing and analysis of overall information required for the decision process of the department of health.

References

[1]A. Berson and S.J.Smith, "Data Warehousing, Data Mining and OLAP", Mc Graw Hill,England, 1999

[2]A Marotta, A transformation based approach for designing Data Warehouse Internal Report . InCo. Universidad dela Republica, Montevideo, Uruguay,1999

[3]C. Surajit and D. Umeshwar, "An Overview of Data Warehousing and OLAP Technology", pp.2-3, Microsoft Research, Redmond, 2000

[4]Galina Bogdanova, Tsvetanka Georgieva, "Analyzing the Data in OLAP Data Cubes"

[5]Jiawei Han, Micheline Kamber , Data Mining: Concepts and Techniques, ISBN 1-55860-489-8, 2001 by Morgan Kanfmann, San Francisco

[6]M.Jarke, M.A. Jeusfeld, C. Quis, P.Vassiliads Architecture and quality in data warehouse. An extended repository Approach Information Systems, 24(3)229-253(1999). A previous version appeared in Proc. 10 the Conf. of Advanced Information Systems Engineering (CASE 98), Pisa, Italy (1998)

[7] W.H Inmon, EIS and the data warehouse: a simple approach to building an effective foundation for EIS. Data base Programming & Design, 5(11): 70-73, November 1992

[8]W.H Inmon, Building the data bridge: the ten critical success factors of building a data warehouse. Data base Programming & Design, 1992

[9]Widom, Research Problems in Data Warehousing, in Proc. 4 th Int. Conf. on Information and Knowledgement, 1995

[10]http://www.cs.ualberta.ca/~zaiane/courses/cmput690/glossary.html

[11]http://www.moh.gov.mm/file/Department %20of%20Health.pdf

[12]http://www.moh.gov.mm/file/Evidence %20for%20Decision.pdf

[13]http://paul.rutgers.edu/~aminabdu/cs541/ cube_op.pdf