Sales System for Golf Course products using Data warehouse and OLAP Technology

May Thynn Kyi, Dr.Nan Si Kham
University of Computer Studies, Yangon
maythynnkyi@gmail.com; nansikham@gmail.com;

Abstract

Many in top managements extremely need high quality information for various time periods. Most organizations are facing some problems concerning data information gap. Decision support data usually needs to be collected from a variety of operational system and kept in a data store of its own on a separate platform.

Data warehouse is constructed by integrating multiple heterogeneous sources and stores historical data, both summarized and detailed information. Online Analytical processing is used to describe the analysis of complex data from the data warehouse. OLAP service has a number of powerful tools that make it the best solution for designing OLAP system. This presents a system that uses multidimensional Online Analytical processing technology for extracting necessary information and knowledge. In this system,, the data warehouse using OLAP query for sale analysis of golf course product data has been developed. The aim of this system is accurate, effective, and efficient access for users by using OLAP query.

Keyword: Data warehouse, OLAP, Multidimensional OLAP.

1. Introduction

Data warehouse is a collection of decision support technologies, aimed at enabling the knowledge worker to make better and faster decisions. Data warehousing technologies have been successfully deployed in many industries, manufacturing, retail, financial services, transportation, telecommunications, utilities and healthcare.

Data warehouse is a subject-oriented, time variant, non volatile collection of data in support of management decision. The data from various sources are extracted, cleaned, loaded and transformed into the data warehouse, and these data are accessed by using query and analysis tools. The data warehouse is structured around the major subject areas. Data in the data warehouse are stored to provide information from a historical perspective. And these data are not updated in real time and it is refreshed from data in operational system on a regular basis.

Most organizations are having a lot of data but poor information and so an information gap, data warehouses which consolidate and integrate information from many different sources and manipulate a meaningful format for making accurate business decision are developed.

Data warehouse generally contain very large amount of data from multiple sources that may include databases from different data models and sometimes files acquired from independent systems and platforms.

2. Related Work

There are many approaches for information retrieval system to provide the decision making. Richard Barker proposed companies set up data warehouses when it is perceived that a body of data is critical to the successful running of their business.[4] Such data may come from a wide variety of sources, and is then typically made available via a coherent database mechanism, such as an Oracle database. By their very nature they tend to be very large in size, used by a large percentage of key employees, and may need to be accessible across the nation, or even the globe [2]. The perceived value of a data warehouse is that executives and others can gain real competitive advantage by having instant access to relevant corporate information. In fact, leading companies now realize that information has become the most potent corporate asset in today's demanding markets.

Since the data warehouse is used for decision making, it is important that data in the warehouse be correct and standardized. The data preparation process includes extracting required data is from various sources, cleansing, transforming and loading these data into the warehouse and then periodical refreshing.

Jiawei Han, Michehine proposed specific dimensions and measures defined for the cubes in any particular OLAP system depend on the kinds of analysis important to the enterprise. [6]

Transforming OLTP data from relational tables into OLAP cubes, and the design of the cubes, is a complex area that is the subject of many third party books.

3. Characteristics of Data Warehousing

Data warehousing is to refer to the characteristics of a data warehouse as follow:

- Subject Oriented
- Integrated
- Non volatile
- Time variant

3.1 Subject-oriented

A data warehouse is organized around major subjects, such as customer, supplier, product, and sales. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

3.2 Integrated

A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and on-line transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on.

3.3 Time-variant

Data are stored to provide information from a historical perspective. (e.g. the past 5-10 years). Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time.

3.4 Nonvolatile

A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation a data warehouse does not require transaction processing, recovery, and concurrency control mechanisms. It usually requires only two operations in data accessing: initial loading of data and access of data.

4. Data Warehouse Architecture

Data warehouses and their architectures vary depending upon the specifics of an organization's situation. Three common architectures are:

- Data Warehouse Architecture (Basic)
- Data Warehouse Architecture (with a Staging Area)
- Data Warehouse Architecture (with a Staging Area and Data Marts)

4.1 Data Warehouse Architecture (with a Staging Area)

This illustrates four things:

- Data Sources (operational file with multiple format from several clinic)
- Staging Area (the process of moving data from the operational database to the data warehouse)
- Data warehouse (meta data, summary data, raw data)
- Users (Analysis, reporting and mining)

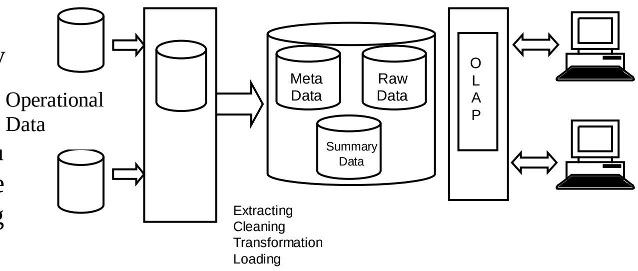


Figure.1 Architecture of a Data Warehouse with Staging Area

5. Data warehouse Process

The process of extracting data from source systems and bring it into data warehouse is commonly called ETL, which stands for extraction, transformation and loading.

5.1 Extraction

Extraction is the operation of extracting data from a source system from future use in a data warehouse environment. This is the first step of the ETL process. After extraction, data can be transformed and loaded into the data warehouse.

Its focus is determining which data needs to be extracted and bring the data into the data warehouse. However, the data sources might be very complex so the designing and creating the extraction is often the most time-consuming task in the ETL process. The data has to be extracted normally not only once, but several times in a periodic manner to supply all changed data into the data warehouse.

5.2 Data Cleaning

Extracting data from remote data sources, especially heterogeneous databases, can bring a lot of error and inconsistent information to the warehouse, i.e. duplicate, alias, and abbreviation and synonym data.

Data cleaning is a process to deal with detecting and removing redundancies and inconsistence from data in order to improve the data quality.

5.3 Transforming

The data sources of a data warehouse contain multiple data structures, while the data warehouse has a single kind of data structure. Each heterogeneous data has to be transformed into a uniform structured data before loading into the data warehouse.

5.4 Loading

Data loading is a process moving data from one data system to another system, especially, make the data become accessible for the data warehouse? Data loading always connects with data transformation in data warehousing.

5.5 Refreshing of Data Warehouse

Refreshing is a process which propagates the updates from the data sources to the warehouse. The warehouse is refreshed periodically, using the transaction systems as its source of data. Note that each data collection has its own refresh cycle governing when its data is periodically updated in the warehouse. For example, one data collection may be refreshed daily, while one other may be refreshed weekly or monthly. Thus, data in the warehouse reflects the most recent refresh cycles for the data collections.

6. Multidimensional Data Model

Data warehouses and OLAP tools are based on a multidimensional data model. This model views data in the form of a data cube.

A data cube allows data to be modeled and views in multiple dimensions. It is defined by dimensions and facts.

In general terms, dimensions are the perspectives or entities with respect to which organization wants to keep records. Each dimension may have a table associated with it, called a dimension table.

A multidimensional data model is typically organized around a central theme. This theme is represented by a fact table. Facts are numerical measures. The fact table contains the names of the facts, or measures, as well as keys to each of the related dimensional tables.

In the data warehousing research literature, a data cube is referred to as cuboids. A lattice of cuboids is showed the data at a different level of summarization, on group by (i.e., summarized by a different subset of the dimensions). The lattice of cuboids is then referred to as a data cube. The cuboids that holds the lowest level of summarization is called the base cuboids. The 0-D cuboid, which holds the highest level of summarization, is called the apex cuboids.

6.1 Schemas for Multidimensional Databases

The entity-relationship data model is commonly used in the design of relational databases, where a database schema consists of a set of entities and the relationships between them. Such a data model is appropriate for on-line transaction processing. A data warehouse, however, requires a concise, subject-oriented schema that facilitates on-line data analysis.

The most popular data model for a data warehouse is a multidimensional model. Such a model can exist in the form of a star schema, snowflake schema or a fact constellation schema.

Star schema: The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table (fact table) containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables (dimension tables), one for each dimension. The schema graph resembles a starburst, with the dimension tables displayed in a radial pattern around the central fact table.

Snowflake schema: The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.

The major difference between the snowflake and star schema models is that the dimension tables of the snowflake model may be kept in normalized form to reduce redundancies.

Fact constellation: Sophisticated applications may require multiple fact tables to share dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation.

6.2 OLAP operations in the Multidimensional Data Model

In the multidimensional model, data are organized into multiple dimensions, and each dimension contains multiple levels of abstraction defined by concept hierarchies. A number of OLAP data cube operations exist to materialize these different views, allowing interactive querying and analysis of the data. OLAP provides a user-friendly environment for interactive data analysis.

Roll-up: The roll-up operation (also called drill-up operations) performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension or by dimension reduction. When roll-up is performed by dimension reduction, one or more dimensions are removed from cube.

Drill-down: Drill-downs is the reverse of roll-up. It navigates from less detailed data to more detailed data. Drill-down can be realized by either stepping down a concept hierarchy for a dimension or introducing additional dimensions. Since a drill-down adds more detail to the data, it can also be performed by adding new dimensions to a cube.

Slice and dice: The slice operation performs a selection on one dimension of the given cube, resulting in a sub cube. The dice operation defines as a sub cube by performing a selection on two or more dimensions.

Pivot (rotate): Pivot (also called rotate) is a visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data.

7. System Implementation

The proposed system is the study of data warehouse methodology and how the various aspects of data warehousing can be supported within the sales analysis of golf course products. By using this system, the user can investigate the sales information quickly without going through travel. This presents a system that uses multidimensional OLAP data mining technology for extracting necessary information and knowledge from golf course product data warehouse. The golf course products include many different types of products. They are flag nylon, dimple tee marker, cast aluminum marker head, pro-ball washer, ball washer stand, Royal line flag stick, riji cup, hole cutter, sprinkler, and so on.

It uses data warehouse with staging area they include four things. They are data source, staging area data warehouse and users. The proposed system uses the star schema. There are one fact table and four dimension table. They are time dimension, location dimension, item type dimension and item dimension. Figure 2 shows the star schema of this system.

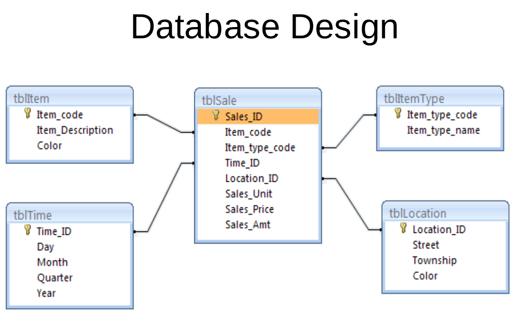


Figure 2. Database Design (Star Schema)

The proposed system includes three parts. They are operational source, data warehouse and OLAP query.

At the first part, daily transactions can be operated on golf course products sales. This part will be used by sale person and operators.

In second part, the data from operational level are loaded into the warehouse through extraction, transformation. Summarized data are prepared in data marts.

At the third part, OLAP (Online Analytical Processing) queries can be executed from data cube and then query, report analysis can be done. This part will be used by manager or decision maker. After analyzing the sales records, the manager can consider and making decision for the promotion, ordering, storing the items on golf course products.

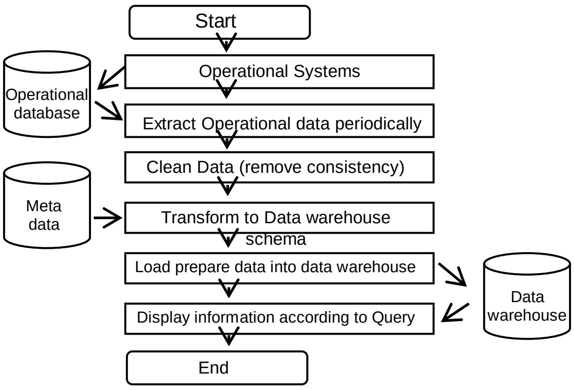


Figure.3 Process Flow Diagram of the system

The manager and decision maker analyses the sales records of golf course product and retrieve summarized data report using online analytical processing (OLAP) tools. It is used to determine the location produces sales together on each of transaction. From the data warehouse of the golf course product information, is by quarterly of the time dimension and by item of item dimension and by township of location dimension. Figure 4 shows the multidimensional data cube of this system.

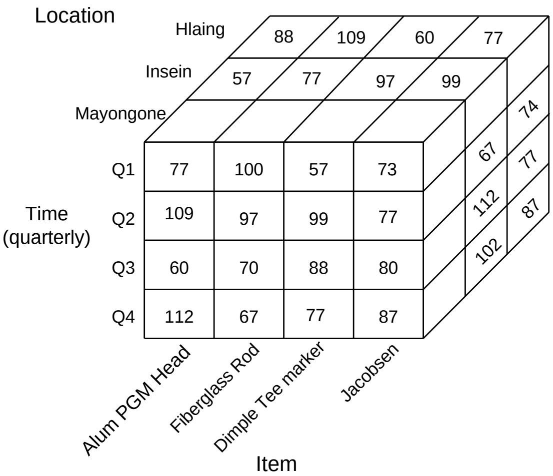
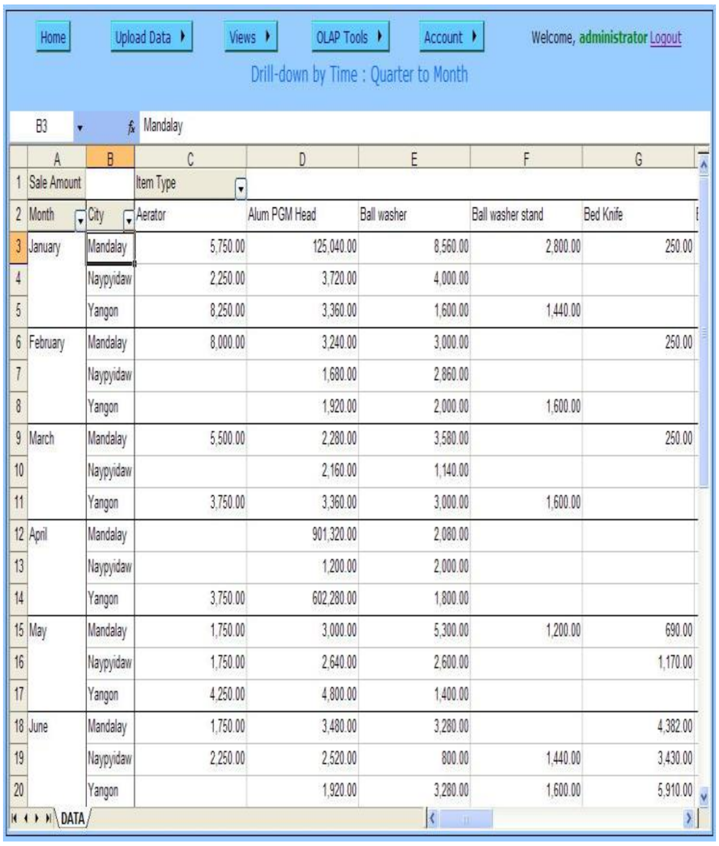


Figure 4. Multidimensional data cube

In the multidimensional model, golf course product data are organized into multiple dimensions, and each dimension contains multiple levels of abstraction defined by concept hierarchies. In this system, all OLAP operations are used. They are roll up, drill down, slice and dice and pivot operations. OLAP provides a user-friendly environment for interactive data analysis.

The user wants to Drill-down operations by time from Quarter to Month in Time attribute, the system shows the drill-down by time form in figure 5.



	A	B	C	D	E	F	G
1	Sale Amount		Item Type				
2	Month	City	Aerator	Alum PGM Head	Ball washer	Ball washer stand	Bed Knife
3	January	Mandalay	5,750.00	125,040.00	8,560.00	2,800.00	250.00
4		Naypyidaw	2,250.00	3,720.00	4,000.00		
5		Yangon	8,250.00	3,360.00	1,600.00	1,440.00	
6	February	Mandalay	8,000.00	3,240.00	3,000.00		250.00
7		Naypyidaw		1,680.00	2,860.00		
8		Yangon		1,920.00	2,000.00	1,600.00	
9	March	Mandalay	5,500.00	2,280.00	3,580.00		250.00
10		Naypyidaw		2,160.00	1,140.00		
11		Yangon	3,750.00	3,360.00	3,000.00	1,600.00	
12	April	Mandalay		901,320.00	2,080.00		
13		Naypyidaw		1,200.00	2,000.00		
14		Yangon	3,750.00	602,280.00	1,800.00		
15	May	Mandalay	1,750.00	3,000.00	5,300.00	1,200.00	690.00
16		Naypyidaw	1,750.00	2,640.00	2,600.00		1,170.00
17		Yangon	4,250.00	4,800.00	1,400.00		
18	June	Mandalay	1,750.00	3,480.00	3,280.00		4,382.00
19		Naypyidaw	2,250.00	2,520.00	800.00	1,440.00	3,430.00
20		Yangon		1,920.00	3,280.00	1,600.00	5,910.00

Figure 5. Drill down operation by time for quarter to month

8. Conclusion

This thesis is the study of data warehouse methodology and how the various aspects of data warehousing can be supported within the sales analysis of imported golf course products. This feature is appropriate for data in warehousing environment, could be used for solving the problem relating to data warehousing activities.

By using this system, the user can investigate the sales information quickly without going through travel. Moreover, the efficient seller get more customers and have more business because of the user can decide when and where to get product for their suitable sales problem by using this system.

References

[1] A.Weininger, Efficient execution of joins in a star schema. In SIGMOD, pages 542-545, 2002.

[2] Berson Alex, j. Smith Stephen, Data warehousing, Data Mining & OLAP, Mc.Graw Hill series on Data warehousing and Data Management, 1997

[3] Chaudhuri, S., Dayal, U: An overview of Data Warehosing and OLAP Technology. ACM

SIGMOD Record 26(1), 1997

[4] Han Jiawei, Kamber Micheline, Data Mining: Concepts and Techniques, ISBN 1-55860-489-8, 2001 by Morgan Kaufmann, San Francisco.
[5] Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P. : Fundamentals of Data Warehouses. Springer, 2000.
[6] Richard Barger, “Managing a Data warehosue”
[7] W.H. Inmon, “ Building the Data Warehouse”
[8] Wikipeida, the free encyclopedia, Data warehouse, http://en.wikipedia. org/wiki/Data warehouse