On-Line Analytical Processing Tools for Land Revenue Management System

Shamsul Sahibuddin & Nor Hawaniah Zakaria
Collaborative Working & Distributed System Group
Faculty of Computer Science & Information System
University Technology Malaysia
{shamsul,hawaniah}@fsksm.utm.my

Abstract- **A functional data warehouse organizes and stores all of the available data needed for informational, analytical processing over a historical time perspective. The whole purpose of building the data warehouse is to give an identified set of users a particular set of data. This study is focusing on data analysis tools for the land revenue system used by a state government. Land Office is one of the major revenue providers for the state government. According to the report provided by the land office in one of the states in 1999, the total income collected through land office was more than RM15 million [6]. A system was developed in 1995 to facilitate the collection process. This system is called *Sistem Pentadbiran Hasil Tanah* (SPHT) or Land Revenue Management System. This paper will describe how a data analytical tool is going to be implemented in the system in order to provide a better mechanism in the decision-making process. New system architecture is proposed in order to maintain the existing system while at the same time providing new analytical processing tools capabilities. Issues regarding the database architecture are also discussed.**

I. INTRODUCTION

In today's world, the competitive edge is contributed less by optimization but more from the proactive use of the information that computer systems have been collecting over the years. Organization is beginning to realize the vast potential of the information that they hold in their establishment. Tapping into this information can significantly improve the quality of their decision making through focus actions.

To make this possible, a response in the form of data warehouse system was initiated. The data warehouse is more than just data; it is also the processes involved in getting the data from table to analysts. The data warehouse provides facility for integrating the data generated in a world of unintegrated information systems [1]. Data warehouses collect information from many sources into a single database [5]. A functional data warehouse organizes and stores all of the available data needed for informational, analytical processing over a historical time perspective.

The whole purpose of building the data warehouse is to give an identified set of users to a particular set of data. There are as many different tools as there are people to use them. These tools can be classified into three broad categories, namely;

- Data dipping tools,
- Data mining tools and
- Data analysis tools [1].

This study is focusing on data analysis tools for the land revenue system use by a state government. This paper will describe how a data analytical tool is implemented in the system in order to provide a better mechanism in decision-making process.

II LAND REVENUE SYSTEM

Land Office is one of the major revenue providers for the state government. According to the report provided by the land office in one of the state in Malaysia, in 1999 the total of income collected through land office was more than RM15 million [6]. A system was developed in 1995 to facilitate the collection process. This system is called *Sistem Pentadbiran Hasil Tanah* (SPHT) or Land Revenue Management System.

SPHT is used to minimize processing backlog and the maintenance of payment record and to reduce paper trail. The main objective of SPHT is to enhance the tax collection process without sacrificing the security aspect while providing easy access to the information. The system was fully operational in 1999. The system allows payment to be made at any district office. The payment transaction occurs daily.

At the end of the month and at the end of the year a number of reports are generated. Among the statements generated are the following:

- PTG-L(10),
- KEW-249,
- LPR-650,
- LPR-D40 and other statements.

These statements are important to the management in making decision and planning for the future of the organizations [6]. These statements provide the overall information on the income collected. With the current system, the management has to rely on the clerical staff in order to get the report. This creates delay in getting

the report to the management and in the process, some on the information becomes outdated.

There exist a number of other weaknesses in SPHT. The system does not provide the correct estimate value, which is used to calculate the collected percentage value. The estimate value used by the current system is based on the nearest value and the highest value that was used in the previous year's calculation. Also this value is static and as a result the organization will not achieve the exact percentage collected value. Another problem exists in the system is when there is a request by the management for the total collected amount. The data has to be extracted and sorted out from the PTG-L(10) statement manually and hence wasting the time and effort of the management [6].

ROLAP tools are SQL-oriented and have tight integration with the relational model. Metadata is use to isolate the user from the underlying complexities. It also differs from the data differs by their range and depth of analytical functionality. It can also be use as data browsers with drill-down capability from aggregation of detailed data [1].

Multidimensional analysis

In order to analyze the data in many dimensions at once, a multidimensional analysis technique is employed. In this technique, matrix arithmetic and sparse matrix optimizations allow the data to be stored efficiently and analyzed rapidly by the loaded dimensions. Allowing the use of multiple separate but

Table 1 : Examples of PTG-L(10) Statements

Kod Hasil (Revenue Code)	*Perihal* (Description)	*Tahun* (Year)	*Anggaran* (Estimate)	*Sehingga Bulan Lalu* (Until Last Month)	*Pungutan Bulan Semasa* (Current Month Collection)	*Jumlah Pungutan* (Total Collection)	*Peratusan Pungutan* (Percentage of Collection)
5612100	Cukai Tanah	1999	5,000,000	2,500,000	1,000,000	3,500,000	70%
	Semasa	2000	5,000,000	3,000,000	1,000,000	4,000,000	80%
5612101	Cukai Tanah Tunggakan	....	....	....	....	....	....
....	....	....	....	....	....	....	....

III DATA ANALYSIS TOOL

Data analysis tools are used to perform complex analysis of data. They normally have a set of analytic functions that allow sophisticated analysis of the data. On-line analytical processing (OLAP) is one example of the tools available. OLAP tools are designed to allow reasonably large quantity of data to be analyzed online.

On-line Analytical Processing (OLAP)

OLAP describes a class of technologies that are designed for live ad-hoc data access technologies. It is about the methods, structures and tools required rapidly accessing and analyzing data. OLAP tools are designed to allow reasonably large quantity of data to be analyzed online. It allows users to quickly perform standard analytical functions on the data and to represent both data and results graphically. OLAP tools can be divided into two categories, which is ROLAP (relational OLAP) and MOLAP (multidimensional OLAP). Both tools differ in the approach used to analyze data, which is the relational approach or the multidimensional approach [1][7][8].

related hypercubes reduces the scarcity of the matrices further.

MOLAP does have an SQL interface in order to allow them to extract data from a relational database. MOLAP tools are used for analyzing aggregated data in conjunction with its dimension data. These aggregations are normally data mart designed specifically for the MOLAP tools. These tools are not recommended for drill down to detailed data at the fact level [1].

IV OLAP FOR LAND REVENUE SYSTEM

An OLAP tools for Land Revenue Management System was proposed based on the weaknesses of current system.

System Architecture

The system architecture for an OLAP is based on Web-based client-server architecture. In a classical client-server architecture, there exist two-tier, namely a client and a server (*Figure 1*). But for a system that is relying on a legacy database and at the same time is trying to take advantage of the world-wide-web infrastructure, then hyper-tier architecture is a better solution

Issues in System Architecture

There are a number of issues that is affecting the design of the OLAP system architecture. Among the issues are as follows:

- *The need to maintain the mainframe-based legacy database.* The investment on the legacy system is enormous. Thus, there is a requirement to maintain the legacy system. The legacy database can also act as the data warehouse for the whole system. The majority of other transactional systems also use the legacy database.
- *The need to take advantage of the world-wide-web infrastructure.* In order to make the system available on different types of platform, there is a need to utilize the world-wide-web. The web will act as a platform of integration with existing environment.
- *The integration of the legacy database with the multi-dimensional database and OLAP server.* The requirement to maintain the existing legacy database necessitates the setting up of a different multi-dimensional database to be used by the OLAP system. The OLAP server not only is required to handle the analytical query, but it also has to communicate with the data warehouse.
- *The integration of the OLAP server with the world-wide-web.* In order to make the OLAP system accessible on the web, the system must be integrated with the web server and web clients. One way to achieve this is by using Java technology.

Hyper-Tier Solutions for OLAP System

The client-server architecture proposed is a hyper-tier approach (*Figure 2*). This architecture is considered the best solution available considering the issues raised in the previous section. In this approach, the back-end server or the fourth-tier is the mainframe, storing the original database. The third-tier is the multi-dimensional database and the OLAP server. The second-tier is the Web server and the Java server connecting the web client with the OLAP system. And at the other end on the first-tier is the web client for the system.

In this proposed architecture, the legacy database on the fourth-tier allows existing transactional system to co-exist. The multi-dimensional database on the third-tier allows the on-line analytical processing to be conducted by the OLAP server. The Web and Java server on the second-tier act as an integrating platform between the OLAP server and the Web client.

Enabling Technologies

The OLAP systems must be implemented on some form of network. There are a number of problems if the system is to be implemented in a local area network or wide area network from scratch. One is the enabling infrastructure in terms of development tools have to be set up [3]. The problem of integration with existing environments and migration is also an issue.

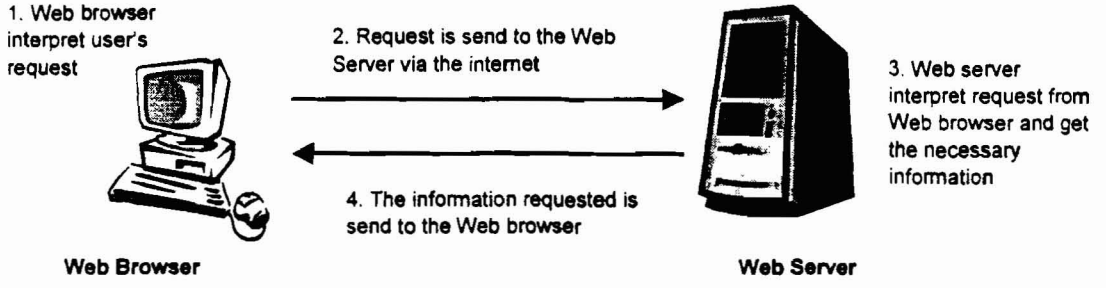

Figure 1: Web-based Client-Server Architecture

Two types of technologies that are of interest to the system architecture are World Wide Web (WWW) and Java Technology (Java). The World Wide Web (WWW) provides a great deal of potential. It can be viewed as a platform of integration with existing environment and as an enabling infrastructure for networked system. The WWW offers simple and unified access for users to retrieve information or to make it available for others (*Figure 1*).

With Java, the issue of migration can be tackled, and provides the tools for extending the WWW and integrating it with existing end-user environment. The Java programming language was designed to meet the challenges of application development in the context of heterogeneous network-wide distributed environment [2]. Java standard classes also provide all the basics blocks necessary for client-server implementation [4]. Java servers allows the request from the Web client to redirected to the OLAP server which in turn will access the multidimensional database.

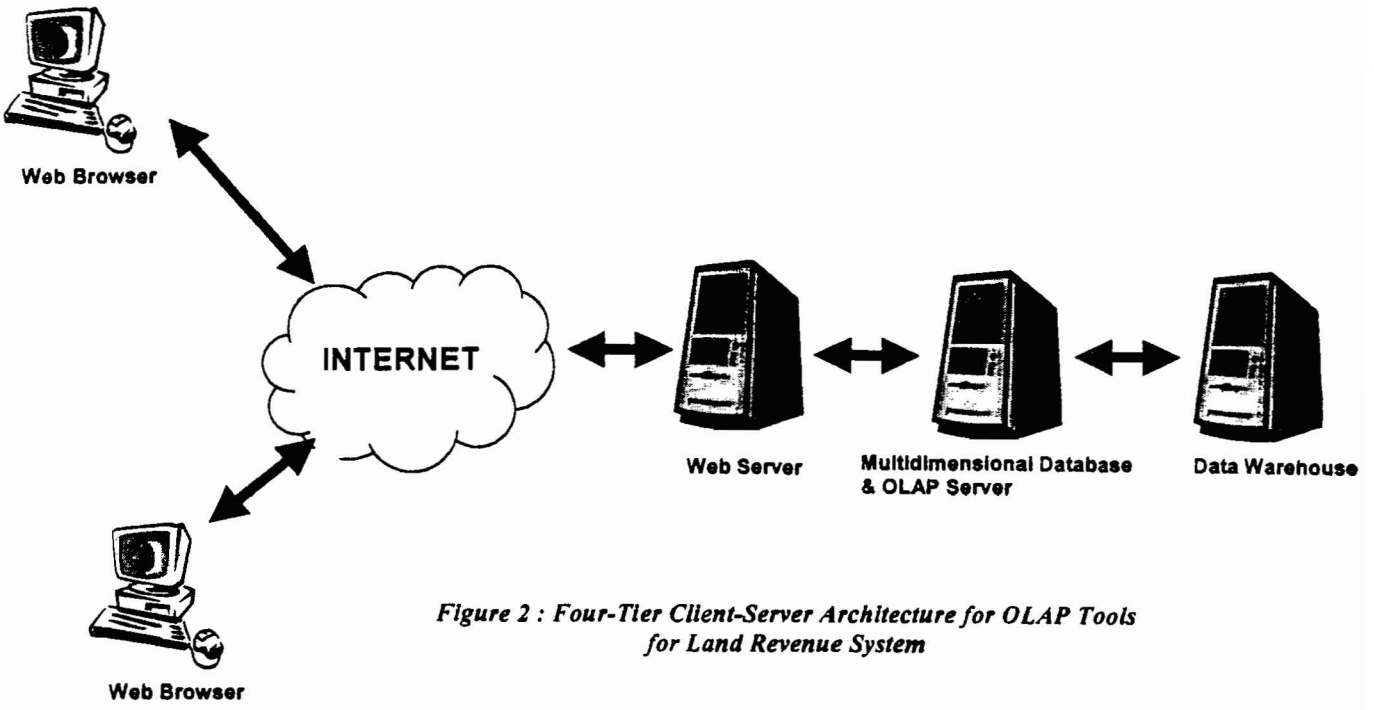

Figure 2 : Four-Tier Client-Server Architecture for OLAP Tools for Land Revenue System

Database Architecture

There are a number of different databases that is involved in OLAP systems. Normally, a separate database is maintained for the operational usage. A different database is also maintained for the data warehouse and the data marts. In this discussion, we are only concern with the database architecture for the multi-dimensional database, which will be used by the OLAP server.

Issues in Database Structure

Organization of data is essential for system performance, especially for OLAP system. One of major technical challenge is to structure a solution that will be effective for a reasonable period of time. The solution offered should not be restructured when business changes or query profiles change.

In designing the database for the OLAP system, several issues shall be focused, among which are:

- *The database schema.* The schema should answer the question of how a database should be designed to answer unknown queries to be performed.
- *The portioning strategy.* The issue here is how the database should be partitioned to allow best performance and manageability of data; and
- *The aggregation (summary) strategy.* This will allow common queries to be performed fast.

Multi-Dimensional Data Model

Arrangement of data is important to allow for multidimensional analysis to happen fast especially in large databases [1][7]. In multidimensional data model, database is viewed as a set of facts in multi-dimensional space. A database schema called star schema describes the physical architecture of the dimensional model.

In order to come-up with a star schema for this OLAP system, there is a need to define the user requirement. It is critical to understand how the information in the data mart will be used. The user requirement will identify the possible analytical functions provided by the systems [8].

In general, the star schema consists of a single fact table and a single table for each dimension. The fact table consists of data that is being used to measure or analyze, and numerical data being tracked, for example the amount of tax paid, or the size of land owned. Whereas, dimension tables are made up of business parameters, for example the type of land info, or the type of tax info. Dimension tables are being set in a denormalized form. The factual data provides the information for analysis.

Several advantages of applying star schema in a multi-dimensional database are as follows:

- Reduce the number of physical joins that normally occur in a normal joined relational tables;
- Simplify the view of the data model;
- Allow relatively easy maintenance on the data

.

V CONCLUSION

As a conclusion, this paper has presented a proposal for an On-line Analytical Processing Tools for Land Revenue System. The system architecture proposed is based on four-tier client-server architecture. The need

to maintain the mainframe-based legacy database, taking advantage of the world-wide-web infrastructure, the integration of the legacy database with the multi-dimensional database and OLAP server, and the integration of the OLAP server with the world-wide-web are the issues that was taken into consideration during the design of the system architecture. The database schema for the multi-dimensional database was also discussed. Star schema, which describes the physical architecture of the dimensional model, is used as the database schema.

REFERENCES

[1] Anahory, Sam and Murray, Dennis, *Data Warehousing in the Real World*, Addison-Wesley: Harlow, England, 1997.

[2] Sahibuddin, Shamsul, "FlexSIG: Flexible Software Inspection Groupware", PhD Thesis, Aston University, Birmingham, United Kingdom, 1999

[3] Doherty, B.S. & Sahibuddin, S., "Modelling Distributed Code Inspection System", *Proceedings of the REDECS'96*, Serdang, Malaysia: Universiti Putra Malaysia Publications, June, pp. 69-73, 1996

[4] Sun, *A Java Language Environment: A White Paper*, Mountain View, California: Sun Microsystems Computer Company, October, 1995

[5] Labio, W.J., Quass,D., & Adelberg, B., *Physical Database Design for Data Warehouses*, http://www.cs.stanford.edu, 20 September, 2000.

[6] Nazri Mohd Nawi, *Pengurusan Hasil Tanah menggunakan Kaedah OLAP*, Laporan Projek I M.Sc., Skudai: FSKSM, Mei, 2000

[7] Pilot Software, *An Introducion to OLAP: Multidimensional Terminology and Technology*, http://www.pilotsw.com/news/olap_white.htm, 21 July, 2000.

[8] Efrem, G.M, *Decision Support and Data Warehouse Systems*, Boston: Irwin McGraw-Hill, 2000

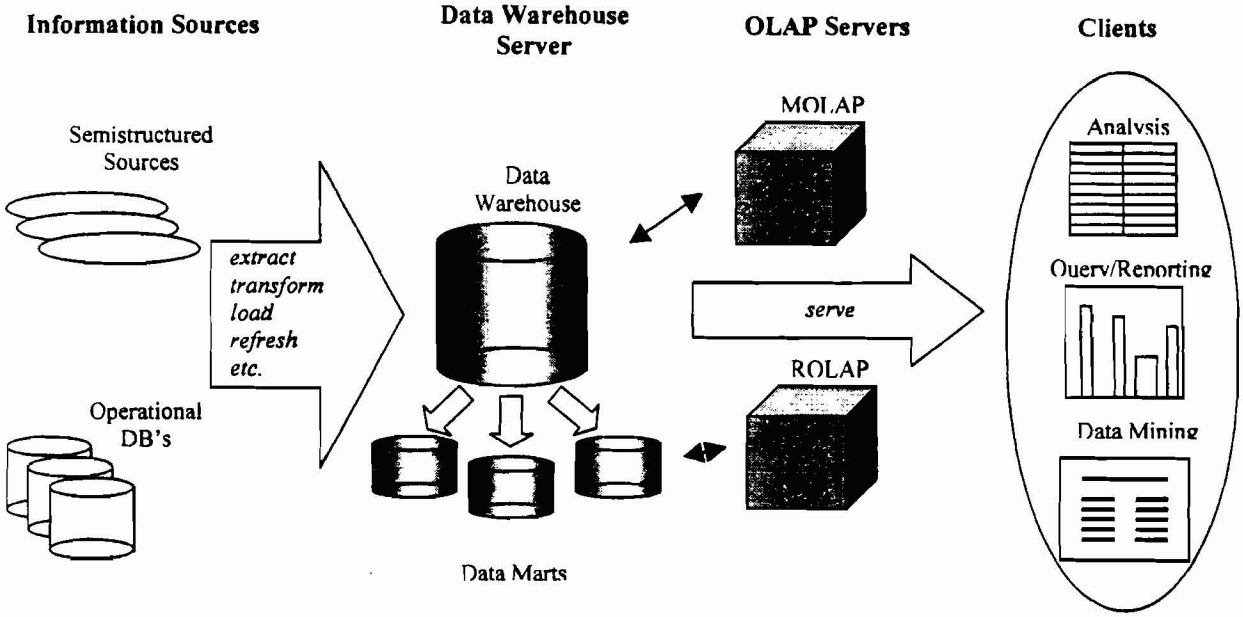

Figure 3 : A Generic Data Warehouse Architecture