A Framework for Healthcare Data Warehousing

Sellappan P.
Department of Information Technology
Malaysia University of Science & Technology
Kelana Square, 47301 Petaling Jaya, Malaysia
sell@must.edu.my

Yih Huey N.
Department of Information Technology
Malaysia University of Science & Technology
Kelana Square, 47301 Petaling Jaya, Malaysia
yhng@must.edu.my

Abstract - *Currently, information collected by the various healthcare providers (hospitals and clinics) in Malaysia is not integrated. Each collects information for its own internal use. So, the information is not aggregated and analyzed at the state or national level to provide useful information for decision making (e.g., by the Ministry of Health). Extracting and aggregating information from heterogeneous and distributed systems have always been a challenge for the healthcare industry. This paper proposes a framework for integrating healthcare information into a single data warehouse. It is simple, cost-effective and yet solves the information integration problem. It consists of a schema mapping process which is partially automated and an Extraction, Transformation and Loading (ETL) process which is fully automated. To demonstrate its viability, a data warehouse containing materialized data is built from several distributed and heterogeneous databases such as SQL Server and Oracle. It is implemented using the Microsoft .NET Framework.*

Keywords: Data Warehouse, Healthcare Information Integration, Extraction Transformation and Loading (ETL), Metamodel, Schema Mapping

1 Introduction

Currently, healthcare providers (i.e., hospitals and clinics) in Malaysia use conventional transaction processing systems to record their daily clinical and financial data. However, as peoples' lifestyles become more complex and technologies become more advanced, users demand more quality or valued-added services. Healthcare providers are expected to increase the value of their transaction processing systems - they need to turn data into actionable information [21].

Healthcare providers generate voluminous data, but they are not leveraged for generating valuable information. Extracting, aggregating and analyzing healthcare information from distributed and heterogeneous data sources (databases) have always been a challenge for the healthcare industry. Aggregating and integrating healthcare information is important for understanding the health of a community because quality services depend on the availability of accurate, relevant and timely information. This paper proposes a healthcare data warehousing that will efficiently integrate healthcare information from distributed and heterogeneous data sources (databases) into a single data warehouse. Such a data warehouse can be used to implement healthcare decision support systems that can enhance the quality of healthcare services provided. It can for example support OnLine Analytical Processing (OLAP) and data mining to generate more focused healthcare information [2].

Most healthcare providers do not use data warehouse technology because it is costly. The framework proposed in this paper solves the data integration problem. It is simple and requires less resources compared to the data integration technologies available in the market. Basically, it provides features to (a) semi-automate the schema mapping process and (b) fully-automate the Extraction, Transformation and Loading (ETL) process. These features allow the integration of healthcare information from heterogeneous and distributed databases.

2 Framework Architecture Design

The framework is designed based on the client-server architecture (Figure 1). It comprises of two main components: Agent and Web Server. The Agent is installed on each client computer. Data extracted from the clients are uploaded to the Web Server for integration via the Internet.

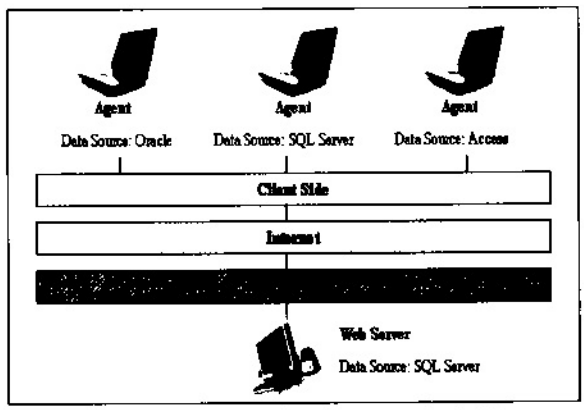

Figure 1 High Level View of the Framework Architecture

The Web Server consists of a data warehouse Manager and an Extraction, Transformation and Loading (ETL) engine (Figure 2). The Manager manages the data warehouse and the metadata. The ETL engine manages the data integration process. A Web application is