The Development of an XML-Based Data Warehouse System

Shi-Ming Huang and Chun-Hao Su

Department of Information Management, National Chung Cheng University
160 San-Hsing, Min-Hsiung, Chia-I 621, Taiwan, R.O.C.
Phone: 886-5-272-0411 ext: 16810
Fax: 886-5-272-3943
smhuang@mis.ccu.edu.tw

Abstract. Along with the enterprise globalization and Internet popularization, the Internet-based Data Warehouse System (DWS) has gradually replaced the traditional DWS and becomes its mainstream structure. The manager can easily obtain and share the data on the distribution system using the Internet. Through the multiple data source collections, the quality and broad base of DWS can be increased and thus help managers to make more decisive policies. However, utilizing the basic client/server structure of DWS can increase many tolerances and cost based problems. This paper uses the XML to establish the Internet-based DWS and utilize the advantage of its flexibility, self-definition, self-description and low cost to improve the unavoidable defect of the client/server DWS. We also use pull and push method approaches to determine what information can be shared on the Internet or delivered through e-mail. In this work, we show that the DWS architecture can not only improve the scalability and speed but also enhance the system security. In addition, it can be applied for both traditional client/server DWS and web-based DWS. We present a case study to prove the validity of this system architecture and create a prototype system to show the feasibility of this system architecture.

1 Introduction

The DWS is the concept of database system developed mainly for the demand to assist the decision making. The background of this development is due to the long time accumulative sources of various exchange data of many companies with the hope using those data to analyze the results for supporting the decision making. In order to observation of various angles of data and make quick changes of analysis, the DWS uses the multi-dimensional data type called Data Cube (DC) [1]. The traditional DWS is built mainly in the closed type network structure of Client/Server system. This kind of structure has the following disadvantages [2].

(1) The two-tier client/server infrastructure makes the system establishment and maintenance difficult and costly. It can only show up the multiple data information by using a special type of the front-end software.

(2) The Information in organization is no longer sufficient for the coming stage of e-commerce, the enterprise Supply Chain Management (SCM) and Customer

Relationship Management (CRM) in one organization must be invested heavily on the related enterprises to make mutual cooperation and to win the customer's confidence.

In today's circumstance, the web-based DWS delivers all of the same kinds of applications that the data warehousing solutions deliver. Nevertheless, a number of challenges still exist in implementing web-based DWS due to immaturity of this technology and to some management concerns. These challenges are Scalability, Speed and Security [3].

In order to solve the challenges of the web-based DWS, this study uses XML structure and adds push and pull technology to form a novel Internet-based DWS. Because XML has more durability and flexibility, self-defined and self-description, highly literary structure and easy usage, we consider that the future Internet-based DWS will be based on XML. With a combination of "pull" and "push" technology, users are presented only with information directly related to their functions, while they still have the option to receive additional information on demand.

2   Related work

2.1   Traditional Client/Server Data Warehouse

Traditionally, the format of the client/server DWS is using the Online Analytical Processing (OLAP) tools to extract the data and to store it temporarily in the data mart. Those data have been reviewed through the procedure of the analysis. The usage data will then show up in many interfaces through the Executive Information System (EIS) [4].

2.2   Web-Based Data Warehouse

The web-based DWS includes client, web server and application server, it allows end users to use web browsers as a user interface in order to access and manipulate data. The server side offers the web page service and controls the flow of information between server and client. The end side of the server side connects to the application side which includes the DWS and the web application format for controlling the storage and extracting the information [2].

2.3   Data Cube and Star Schema

A DC is a multi-dimensional database. It includes a collection of at least one fact table and a set of dimension tables. The purpose of the DWS is to design analytically for answering various queries. Using the method of the DC, the system can quickly respond and satisfy the user's demand.

Due to the fact that the data is stored in the relational DWS, the original data after analysis, conversion and filtering will be stored in the star schema and becomes the data cube. Therefore each DC has its own characteristics. One can often inquire and