Data Warehouse Software

Huanyu Ouyang
People's Hospital of Jangxi Province, China

John Wang
Montclair State University, USA

INTRODUCTION

A *data warehouse* (DW) is a complete intelligent data storage and information delivery or distribution solution enabling users to customize the flow of information through their organization (Inmon & Hackathorn, 2002). It provides all authorized members of users' organization with flexible, secure, and rapid access to critical information and intelligent reporting. DW can extract information from sources anywhere in the world and then delivers intelligence anywhere in the world. It connects to any platform, database, data source, and it will also scale to businesses and applications of any size. As early as the 1970's, data warehousing software (DWS) was recognized when the earliest systems were first developed. The database designs of operational systems were not effective enough for the information analysis and reporting (The Data Warehousing Information Center, 2006).

Today, many corporations are experiencing significant business benefits by using DWS technology. DWS is a separate architecture used to maintain critical historical data that has been extracted from operation data storage and transformed into formats understandable to the organization's analytical community. DWS is a system for storing, retrieving, and managing large amounts of any type of data, and it is a copy of transaction data specifically structured for query and analysis. It is also a complete, powerful, scalable, and customizable intelligent DW solution, which also optionally offers the most complete analytic functionality available on the market, fully-integrated into the system.

BACKGROUND

DW is a field that has grown out of the integration of a number of different technologies and experiences over the last two decades. These experiences have allowed the IT industry to identify the key problems that have to be solved (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 2000). According to Bill Inmon (2001), known as the father of DW, a DW is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions. Subject-oriented means that all relevant data about a subject is gathered and stored as a single set in a useful format. Integrated refers to data being stored in a globally-accepted fashion with consistent naming conventions, measurements, encoding structures, and physical attributes, even when the underlying operational systems store the data differently. Non-volatile means the DW is read-only: data are loaded into the DW and accessed there. Time-variant data represent long-term data, from 5 to 10 years as opposed to the 30- to 60-day time periods of operational data.

DWs build on a database and database schema customized for user's particular business. The solution can be installed either inclusive of a high performance database engine or as a database schema compatible with most industry standard databases. So, it will seamlessly integrate into existing database systems. DW does not depend on one particular database vendor or hardware platform, it is itself entirely platform-independent, and the main DW will connect to any database format, and hence can efficiently combine and pool information from multiple sources. The software will run on servers with multiple processors, or banks of multiple-processor servers for super-computer like performance. The system will scale effortlessly and economically to even huge data sizes and analysis problems. Most of DWs are used for post-decision monitoring of the effects of decisions or for operational issues (Inmon, Welch, & Glassey, 2000).

MAIN FOCUS

Data management needs for large companies evolve at alarming rates. On average, businesses double their volume of data annually. As a result, it is expected that DW will actually be measuring in petabytes instead of terabytes in the next several years. As transaction processing time and synchronization of business functions become ever more essential steps to serving the customer, businesses look for more efficient business intelligence solutions. DWS can often provide such solutions, and its general popularity in major industries has transformed data management. Not only has it fueled competition in business by providing a means to do things faster, better, and for less than the competitors, but DWS in itself has become a competitive industry. Creating the most compatible, user-friendly solutions in a software package is an unending battle among DWS producers.

Benefits and Drawbacks of Data Warehousing

The many advantages to using a DW are helping end-user access to a wide variety of data, increase data consistency, and productivity while decreasing computing costs. It is also able to bring together data from many different sources in one place and provides an infrastructure that could support changes to data and replication of the changed data back into the operational systems. Another very important advantage that the DW has over the general operational systems is its accessibility through a variety of modern software tools. A lot of software can be distinguished on the basis of how suitable it is for the task at hand or how recognizable it is to the customer.

Before the arrival of DWS, problems such as heterogeneous data, legacy data, data unsaved in *online transaction processing* (OLTP) databases, and files made primarily for small, conventional transactions, often signified unbearable delays in the production of reports usually needed for appropriate decision making.

It is a departure from the existing model for operational systems where the primary availability is through custom applications programs written to succeed in specific tasks. For some companies, they view DWS as complex, costly, and almost certain to fail. But more companies are successfully building and operating DWs that boost operational efficiency, lower operating costs, and get them closer to their customers. Information Week (2006) listed the following benefits: (1) merging subject specific data together to create information; (2) standardizing data across the organization; (3) improving turnaround time for reporting; (4) lowering costs to print and distribute reports; and (5) sharing data or allowing others to easily access your data.

Some of the disadvantages are extracting, cleaning, and loading data could be time consuming. A DW project scope might increase, and there may be problems with compatibility with systems already in place. Security could develop into a serious issue; especially if the DW is web accessible (DMReview (B), 2006).

Let's look at why a company would want to build a DW in the first place. Seemingly, many companies have information needs that are currently not being satisfied by their operational and decision support systems. In all possibility, the company is suffering from considerable losses due to the non-availability of this important information. These losses can range from anything such as losing a business to the competition or losing customers to cost overruns. Most likely, the reason for not being able to help satisfy their information needs is due to dirty data, redundant data, inaccessible data, untimely data, and so forth. In other words, companies can suffer real business pain, which they cannot solve with their current systems. This pain usually translates into high losses. When companies go over the cost of a DW solution (which addresses all the data issues when finished accurately), the company has to decide if the benefits (profits to be gained or losses to be eliminated by having their information needs satisfied) overrides the costs of building a DW. If it doesn't, a DW solution is not suitable and maybe it should not be built.

Another disadvantage of DWS is when it is poorly designed. Indeed, a DW will certainly introduce data quality issues as well as many other processes, data, and resource changes—however, all for the better. It is important to understand that the maintenance of a DW can be a time-consuming and difficult task. A company must make the commitment to supply a dedicated support team, appropriate hardware and software, and business sponsorship to make it effective.