Data Mining and Business Intelligence Dashboards

Bhavana Jamalpur, SR Engineering College, Warangal, India
S. S. V. N Sharma, Vaagdevi Engineering College, Warangal, India

ABSTRACT

Today, the most successful companies are those that can respond quickly and flexibly to market changes and opportunities. A detailed understanding and knowledge is required to analyze the business environment at different levels. This can be achieved by implementing value added information throughout the organization. There has been an increased demand in the development of IT that performs data manipulation operations to report and analyze the data. While sifting through all that data, organizing it, and then using it to make the business run more profitably has become a bigger challenge. Decision markers can use such information to implement business strategies that best meet the needs of the ultimate customer. Making use of that information is all the more urgent.

Keywords: Affinity, Business Intelligence, Dashboards, Data Mining, Datawarehousing, Knowledge Discovery in Databases (KDD), Online Analytical Processing (OLAP)

INTRODUCTION

Analytical processing basically can be done in two ways. One is to work directly with the operational system using software tools and components known as front-end tools and middleware. This option can be optimal for companies that do not have a large number of end users running queries and conducting analyses against the operating systems. Since the mid-1990s, a wave of front-end tools that allow end users to conduct queries and report on data stored in operational databases have become available. The problem with this approach, however, is that the tools are effective only with end users who have a medium- to high-degree of knowledge about databases. These limitations call for a second, improved option of analytical processing, which involves three concepts:

1. A business representation of data for end users;
2. A user-friendly, Web-based environment that gives the customers and corporate employees query and reporting capabilities;
3. A single, server-based data repository—a data warehouse (DW)—that allows centralized analysis, security, and control over the data.

DOI: 10.4018/jabim.2012100104
DATA WAREHOUSES

The purpose of a data warehouse is to establish a repository that makes operational data accessible in a form readily acceptable for analytical processing activities, such as EC applications, decision support, and other end-user applications. As part of this accessibility, detail-level operational data must be transformed into a relational form, which makes them more amenable to analytical processing. Thus, data warehousing is not a concept by itself, but is interrelated with data access, retrieval, analysis, and visualization. Not all data are transferred to the data warehouse, and frequently only a summary of the data is transferred in a process of extraction, transformation, and load (ETL). The data that are transferred are organized within the warehouse as a relational database so that it is easy for end users to access. Also, the data are organized by subject, such as by product, customer segment, or business partner. EC data also may be organized according to a business process, such as ordering, shipping, or available inventory. The data then can be optionally replicated in data marts.

DATA ANALYSIS AND KNOWLEDGE DISCOVERY

Once the data are in the data warehouse and/or data marts, they can be accessed by end users. Users can then conduct several types of analytical activities with the data, ranging from decision support and executive support analyses to ad-hoc queries, online analytical processing (OLAP), and data mining.

Operational data store: A database for use in transaction processing (operational) systems that uses data warehouse concepts to provide clean data.

Knowledge Discovery

The process of extracting useful knowledge from volumes of data is known as knowledge discovery in databases (KDD), or just knowledge discovery (KD). The objective of KDD is to identify valid, novel, potentially useful, and ultimately understandable patterns in data. KDD is useful because it is supported by three technologies that are now sufficiently mature to produce meaningful data, massive data collection, powerful multiprocessor computers, and data mining algorithms. Formal computer-based knowledge discovery has been done since the 1960s. However, the enabling techniques have been expanded and improved over time. KDD processes have appeared under various names and have shown different.

DATA MINING

Data mining is the extraction of predictive information from large databases. Recently, there has been a growing emphasis on exploratory analysis of very large datasets to discover useful patterns and/or correlations among attributes. Data mining helps customers leverage their data more effectively and obtain insightful information that can give them a competitive edge. Simply put, data mining software enables customers to discover previously undetected facts present in their business-critical data -- data that may consume many gigabytes or terabytes of storage, may reside in files or various DBMS-managed databases, and may be stored on a variety of operating system platforms. Accuracy, efficiency, and an open architecture are important requirements of such data mining software.

Data Mining: Verification vs. Discovery

Decision support systems (DSS), executive information systems, and query/report writing tools are used to produce reports about data, usually aggregating it through any number of dimensions. Another use of these tools is to detect trends and patterns in customer data that will help answer some questions about the business. When used in this mode, a query is created to access the records relevant to the question(s) being formulated. After the data is retrieved, it is examined to detect the existence of patterns.