Modeling-Centered Data Warehousing Learning: Methods, Concepts and Resources

Nenad Jukic, Quinlan School of Business, Loyola University Chicago, Chicago, IL, USA
Boris Jukic, Operations & Information Systems, School of Business, Clarkson University, Potsdam, NY, USA

ABSTRACT

Though data warehousing is widely recognized in the industry as the principal decision support system architecture and an integral part of the corporate information system, the majority of academic institutions in the US and worldwide have been slow in developing curriculums that reflect this. The authors examine the issues that have contributed to the lag in the coverage of data warehousing topics at universities and introduce methods, concepts and resources that can enable business educators to deal with these issues and conduct comprehensive, detailed, and meaningful coverage of data warehouse related topics.

Keywords: Business Education, Data Warehousing, Databases, Decision Support, Dimensional Modeling, Entry-Relationship Modeling

INTRODUCTION

The corporate data warehouse has become the principal decision support system architecture for most large companies worldwide. Even though data warehousing is widely recognized in the industry as an integral part of the corporate information system infrastructure, the majority of the academic institutions in the US and worldwide have been slow in developing curriculums that reflect this reality. Since the advent of data warehousing in the 1990’s, industry seminars, vendor specific tutorials, and on-the-job training have been the exclusive means of education for the majority of people involved in corporate data warehousing projects. In many cases, the lack of formal academic education related to data warehousing has left these information systems professionals without a clear and meaningful understanding of the overall purpose of the data warehousing process and its various stages. This has been one of the contributing factors to the failure rates of data warehousing projects, which are by some estimates higher than 50% (Frolick & Lindsey, 2003; Hwang & Xu, 2007).

Our goal is to bring the attention of the information systems academic community to this issue by examining the data warehouse academic-education challenges and offering concrete solutions. This paper presents teaching and learning methods, concepts and resources that facilitate introducing the area of...
data warehousing in an academic setting in a way that provides a consequential and lasting understanding of this important area of modern information technology. At the same time, we show how these methods, concepts and resources can be introduced into the business information system curriculum in a swift and straightforward manner.

This paper is organized as follows. First, we give a brief overview of data warehousing and related issues. Afterwards a description of contemporary challenges and approaches related to data warehousing teaching and learning. Then a section that describes methods, concepts, and resources that we developed for facilitating data warehouse-related education. Finally, the last section offers a brief summary and conclusion.

BACKGROUND OVERVIEW-DATA WAREHOUSING

A typical organization maintains and utilizes a number of operational databases. These operational databases are used to support the organization’s day-to-day operations. A data warehouse is created within an organization as a separate database, (using its own DBMS), whose primary purpose is data analysis for the support of management’s decision making processes (Inmon, 2005). The data stored in the data warehouse captures many different aspects of the business process, such as production, supply-chain management, marketing, sales and customer service. This data reflects strategically important information such as customer behavior patterns, sales trends, outcomes of marketing strategies, and other characteristics. Therefore, this data is of vital importance to the success of the business whose state it captures. That is the reason why companies choose to engage in the relatively expensive and lengthy undertaking of creating and maintaining the data warehouse, often containing multiple terabytes of data. For example, one study (Gray, 2006) reports a typical cost of $3 million for creating a 1-terabyte data warehouse, with a typical implementation time of 2 years.

Often the same data can have both operational and analytical purposes, and subsequently can be stored in both an operational database and the data warehouse. For example, data describing that product A was bought by customer B in store C can be stored in an operational data store for business-process support purposes, such as financial transaction record keeping or inventory monitoring. That same piece of information can also be stored in a data warehouse where, combined with vast amounts of similar data accumulated over a time period, it is used to analyze important trends, such as sales patterns or customer behavior.

Why should the same information be stored in two separate places? There are two main reasons that necessitate the creation of a data warehouse as a separate analytical data store. The first reason is the performance of queries in different contexts. Operational queries are mostly short and fast, while analytical queries are complex and consume a significant amount of time. The performance of operational queries can be severely diminished if they have to compete with analytical queries for computing resources. The second reason lies in the fact that, even if performance is not an issue, it is often impossible to structure a database which can be used in a straightforward manner for both operational and analytical purposes. Therefore, a data warehouse is created as a separate data store, designed for accommodating analytical queries. A typical data warehouse periodically retrieves selected analytically-useful data from the operational data sources. The process and the infrastructure that facilitate the retrieval of the data from the operational databases into the data warehouses are known as ETL, which stands for Extraction, Transformation and Load. Figure 1 illustrates a layout of a typical data warehouse.

A data mart is a data store based on the same principles as a data warehouse, but with a more limited scope. Whereas a data warehouse
Class-Based Weighted NB for Text Categorization
www.igi-global.com/chapter/class-based-weighted-nb-for-text-categorization/107249?camid=4v1a