

Chapter 2.13

Development and Design Methodologies in DWM

James Yao

Montclair State University, USA

John Wang

Montclair State University, USA

Qiyang Chen

Montclair State University, USA

June Lu

University of Houston – Victoria, USA

INTRODUCTION

Information systems were developed in early 1960s to process orders, billings, inventory controls, payrolls, and accounts payables. Soon information systems research began. Harry Stern started the “Information Systems in Management Science” column in *Management Science* journal to provide a forum for discussion beyond just research papers (Banker & Kauffman, 2004). Ackoff (1967) led the earliest research on management information systems for decision-making purposes and published it in *Management Science*. Gorry and Scott Morton (1971) first used the term *decision support systems* (DSS) in a paper and constructed a framework for improving management

information systems. The topics on information systems and DSS research diversifies. One of the major topics has been on how to get systems design right.

As an active component of DSS, data warehousing became one of the most important developments in the information systems field during the mid-to-late 1990s. It has been estimated that about 95% of the *Fortune 1000* companies either have a data warehouse in place or are planning to develop one (Wixon & Watson, 2001). Data warehousing is a product of business need and technological advances. Since business environment has become more global, competitive, complex, and volatile customer relationship management (CRM) and e-commerce initiatives are creating requirements for large, integrated data repositories and advanced analytical capabilities.

DOI: 10.4018/978-1-59904-843-7.ch029

By using a data warehouse, companies can make decisions about customer-specific strategies such as customer profiling, customer segmentation, and cross-selling analysis (Cunningham, Song, & Chen, 2006). To analyze these large quantities of data, data mining has been widely used to find hidden patterns in the data and even discover knowledge from the collected data. Thus how to design and develop a data warehouse and how to use data mining in the data warehouse development have become important issues for information systems designers and developers.

This article presents some of the currently discussed development and design methodologies in data warehousing and data mining, such as the multidimensional model vs. relational entity-relationship (ER) model, corporate information factory (CIF) vs. multidimensional methodologies, data-driven vs. metric-driven approaches, top-down vs. bottom-up design approaches, data partitioning and parallel processing, materialized view, data mining, and knowledge discovery in database (KDD).

BACKGROUND

Data warehouse design is a lengthy, time-consuming, and costly process. Any wrongly calculated step can lead to a failure. Therefore, researchers have placed important efforts to the study of design and development related issues and methodologies.

Data modeling for a data warehouse is different from operational database, for example, online transaction processing (OLTP), data modeling. An operational system is a system that is used to run a business in real time, based on current data. An OLTP system usually adopts ER modeling and application-oriented database design (Han & Kamber, 2006). An information system, like a data warehouse, is designed to support decision making based on historical point-in-time and prediction data for complex queries or data min-

ing applications (Hoffer, Prescott, & McFadden, 2007). A data warehouse schema is viewed as a dimensional model (Ahmad, Azhar, & Lukauskis, 2004; Han & Kamber, 2006; Levene & Loizou, 2003). It typically adopts either a star or snowflake schema and a subject-oriented database design (Han & Kamber, 2006). The schema design is the most critical to the design of a data warehouse.

Many approaches and methodologies have been proposed in the design and development of data warehouses. Two major data warehouse design methodologies have been paid more attention. Inmon, Terdeman, and Imhoff (2000) proposed the CIF architecture. This architecture, in the design of the atomic-level data marts, uses denormalized entity-relationship diagram (ERD) schema. Kimball (1996, 1997) proposed multidimensional (MD) architecture. This architecture uses star schema at atomic-level data marts. Which architecture should an enterprise follow? Is one better than the other? Currently, the most popular data model for data warehouse design is the dimensional model (Bellatreche & Mohania, 2006; Han & Kamber, 2006). Some researchers call this model the data-driven design model. Artz (2006) advocates the metric-driven view, which, as another view of data warehouse design, begins by identifying key business processes that need to be measured and tracked over time in order for the organization to function more efficiently. There has always been the issue of top-down vs. bottom-up approaches in the design of information systems. The same is with a data warehouse design. These have been puzzling questions for business intelligent architects and data warehouse designers and developers. The next section will extend the discussion on issues related to data warehouse and mining design and development methodologies.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/development-design-methodologies-dwm/48563

Related Content

A Theoretical Framework for a Simulation-Based Analysis of Supply Chain Risk Management

Ruslan Klimov, Yuri Merkurjev and Juri Tolujew (2010). *Managing Risk in Virtual Enterprise Networks: Implementing Supply Chain Principles* (pp. 162-183).

www.irma-international.org/chapter/theoretical-framework-simulation-based-analysis/42220

Exploring the Determinants of Intent to Adopt Mobile Value-Added Services by Women's Micro and Small Businesses in Tonga

Malia Benedine Faasolo and Eli Sumarliah (2021). *International Journal of Enterprise Information Systems* (pp. 16-36).

www.irma-international.org/article/exploring-the-determinants-of-intent-to-adopt-mobile-value-added-services-by-womens-micro-and-small-businesses-in-tonga/289843

Assessing the Impact of System Quality, Information Quality, and Service Quality on Enterprise Resource Planning (ERP) Systems

Godwin Banafo Akrong, Yunfei Shao and Ebenezer Owusu (2021). *International Journal of Enterprise Information Systems* (pp. 69-84).

www.irma-international.org/article/assessing-the-impact-of-system-quality-information-quality-and-service-quality-on-enterprise-resource-planning-erp-systems/289845

Monitoring Organizational Transactions in Enterprise Information Systems with Continuous Assurance Requirements

Rui Pedro Marques, Henrique Santos and Carlos Santos (2015). *International Journal of Enterprise Information Systems* (pp. 13-32).

www.irma-international.org/article/monitoring-organizational-transactions-in-enterprise-information-systems-with-continuous-assurance-requirements/124782

Success Factors and Performance Indicators for Enterprise Application Integration

A. Schwinn (2007). *Enterprise Architecture and Integration: Methods, Implementation and Technologies* (pp. 23-39).

www.irma-international.org/chapter/success-factors-performance-indicators-enterprise/18360