

# Administering and Managing a Data Warehouse

**James E. Yao**

*Montclair State University, USA*

**Chang Liu**

*Northern Illinois University, USA*

**Qiyang Chen**

*Montclair State University, USA*

**June Lu**

*University of Houston-Victoria, USA*

## INTRODUCTION

As internal and external demands on information from managers are increasing rapidly, especially the information that is processed to serve managers' specific needs, regular databases and decision support systems (DSS) cannot provide the information needed. Data warehouses came into existence to meet these needs, consolidating and integrating information from many internal and external sources and arranging it in a meaningful format for making accurate business decisions (Martin, 1997). In the past five years, there has been a significant growth in data warehousing (Hoffer, Prescott, & McFadden, 2005). Correspondingly, this occurrence has brought up the issue of data warehouse administration and management. Data warehousing has been increasingly recognized as an effective tool for organizations to transform data into useful information for strategic decision-making. To achieve competitive advantages via data warehousing, data warehouse management is crucial (Ma, Chou, & Yen, 2000).

## BACKGROUND

Since the advent of computer storage technology and higher level programming languages (Inmon, 2002), organizations, especially larger organizations, have put enormous amount of investment in their information system infrastructures. In a 2003 IT spending survey, 45% of American company participants indicated that their 2003 IT purchasing budgets had increased compared with their budgets in 2002. Among the respondents, database applications ranked top in areas of technology being implemented or had been implemented, with 42% indicating a recent implementation (Informa-

tion, 2004). The fast growth of databases enables companies to capture and store a great deal of business operation data and other business-related data. The data that are stored in the databases, either historical or operational, have been considered corporate resources and an asset that must be managed and used effectively to serve the corporate business for competitive advantages.

A database is a computer structure that houses a self-describing collection of related data (Kroenke, 2004; Rob & Coronel, 2004). This type of data is primitive, detailed, and used for day-to-day operation. The data in a warehouse is derived, meaning it is integrated, subject-oriented, time-variant, and nonvolatile (Inmon, 2002). A data warehouse is defined as an integrated decision support database whose content is derived from various operational databases (Hoffer, Prescott, & McFadden, 2005; Sen & Jacob, 1998). Often a data warehouse can be referred to as a multidimensional database because each occurrence of the subject is referenced by an occurrence of each of several dimensions or characteristics of the subject (Gillenson, 2005). Some multidimensional databases operate on a technological foundation optimal for "slicing and dicing" the data, where data can be thought of as existing in multidimensional cubes (Inmon, 2002). Regular databases load data in two-dimensional tables. A data warehouse can use OLAP (online analytical processing) to provide users with multidimensional views of their data, which can be visually represented as a cube for three dimensions (Senn, 2004).

With the host of differences between a database for day-to-day operation and a data warehouse for supporting management decision-making process, the administration and management of a data warehouse is of course far from similar. For instance, a data warehouse team requires someone who does routine data extraction, transformation, and loading (ETL) from operational data-

bases into data warehouse databases. Thus the team requires a technical role called ETL Specialist. On the other hand, a data warehouse is intended to support the business decision-making process. Someone like a business analyst is also needed to ensure that business information requirements are crossed to the data warehouse development. Data in the data warehouse can be very sensitive and cross functional areas, such as personal medical records and salary information. Therefore, a higher level of security on the data is needed. Encrypting the sensitive data in data warehouse is a potential solution. Issues as such in data warehouse administration and management need to be defined and discussed.

## **MAIN THRUST**

Data warehouse administration and management covers a wide range of fields. This article focuses only on data warehouse and business strategy, data warehouse development life cycle, data warehouse team, process management, and security management to present the current concerns and issues in data warehouse administration and management.

## **Data Warehouse and Business Strategy**

“Data is the blood of an organization. Without data, the corporation has no idea where it stands and where it will go” (Ferdinandi, 1999, p. xi). With data warehousing, today’s corporations can collect and house large volumes of data. Does the size of data volume simply guarantee you a success in your business? Does it mean that the more data you have the more strategic advantages you have over your competitors? Not necessarily. There is no predetermined formula that can turn your information into competitive advantages (Inmon, Terdeman, & Imhoff, 2000). Thus, top management and data administration team are confronted with the question of how to convert corporate information into competitive advantages.

A well-managed data warehouse can assist a corporation in its strategy to gain competitive advantages. This can be achieved by using an exploration warehouse, which is a direct product of data warehouse, to identify environmental factors, formulate strategic plans, and determine business specific objectives:

- **Identifying Environmental Factors:** Quantified analysis can be used for identifying a corporation’s products and services, market share of specific products and services, financial management.

- **Formulating Strategic Plans:** Environmental factors can be matched up against the strategic plan by identifying current market positioning, financial goals, and opportunities.
- **Determining Specific Objectives:** Exploration warehouse can be used to find patterns; if found, these patterns are then compared with patterns discovered previously to optimize corporate objectives (Inmon, Terdeman, & Imhoff, 2000).

While managing a data warehouse for business strategy, what needs to be taken into consideration is the difference between companies. No one formula fits every organization. Avoid using so called “templates” from other companies. The data warehouse is used for your company’s competitive advantages. You need to follow your company’s user information requirements for strategic advantages.

## **Data Warehouse Development Cycle**

Data warehouse system development phases are similar to the phases in the systems development life cycle (SDLC) (Adelman & Rehm, 2003). However, Barker (1998) thinks that there are some differences between the two due to the unique functional and operational features of a data warehouse. As business and information requirements change, new corporate information models evolve and are synthesized into the data warehouse in the Synthesis of Model phase. These models are then used to exploit the data warehouse in the Exploit phase. The data warehouse is updated with new data using appropriate updating strategies and linked to various data sources.

Inmon (2002) sees system development for data warehouse environment as almost exactly the opposite of the traditional SDLC. He thinks that traditional SDLC is concerned with and supports primarily the operational environment. The data warehouse operates under a very different life cycle called “CLDS” (the reverse of the SDLC). The CLDS is a classic data-driven development life cycle, but the SDLC is a classic requirements-driven development life cycle.

## **The Data Warehouse Team**

Building a data warehouse is a large system development process. Participants of data warehouse development can range from a data warehouse administrator (DWA) (Hoffer, Prescott, & McFadden, 2005) to a business analyst (Ferdinandi, 1999). The data warehouse team is supposed to lead the organization into assuming their roles and thereby bringing about a part-

4 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/administering-managing-data-warehouse/10558](http://www.igi-global.com/chapter/administering-managing-data-warehouse/10558)

## Related Content

---

### Metadata- and Ontology-Based Semantic Web Mining

Marie Aude Aufaure, Bénédicte Le Grand, Michel Sotoand Nacera Bennacer (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 3531-3556).

[www.irma-international.org/chapter/metadata-ontology-based-semantic-web/7848](http://www.irma-international.org/chapter/metadata-ontology-based-semantic-web/7848)

### Expanding Data Mining Power with System Dynamics

Edilberto Casado (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2688-2696).

[www.irma-international.org/chapter/expanding-data-mining-power-system/7792](http://www.irma-international.org/chapter/expanding-data-mining-power-system/7792)

### Modeling Web-Based Data in a Data Warehouse

Hadrian Peterand Charles Greenidge (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 826-831).

[www.irma-international.org/chapter/modeling-web-based-data-data/10711](http://www.irma-international.org/chapter/modeling-web-based-data-data/10711)

### Online Signature Recognition

Indrani Chakravarty, Nilesh Mishra, Mayank Vatsa, Richa Singhand P. Gupta (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 885-890).

[www.irma-international.org/chapter/online-signature-recognition/10721](http://www.irma-international.org/chapter/online-signature-recognition/10721)

### Swarm Quant' Intelligence for Optimizing Multi-Node OLAP Systems

Jorge Loureiroand Orlando Belo (2009). *Progressive Methods in Data Warehousing and Business Intelligence: Concepts and Competitive Analytics* (pp. 132-154).

[www.irma-international.org/chapter/swarm-quant-intelligence-optimizing-multi/28165](http://www.irma-international.org/chapter/swarm-quant-intelligence-optimizing-multi/28165)