A Metadata Oriented Architecture for Building Datawarehouse

HEESEOK LEE, TAEHUN KIM, AND JONGHO KIM
Korea Advanced Institute of Science and Technology, Korea

Data warehouse is an intelligent store of data that can aggregate vast amounts of information. A metadata is critical for implementing data warehouse. Therefore, integrating data warehouse with its metadata offers a new opportunity to create a more adaptive information system. This paper proposes a metadata-oriented data warehouse architecture that consists of seven components: legacy system, extracting software, operational data store, data warehouse, data mart, application, and metadata. A taxonomy for dataflow and metaflow is proposed for better understanding of the architecture. In addition, a metadata schema is built within the framework of the seven components. The architecture with its metadata component is applied to a real-life data warehouse for a large medical center in order to illustrate its practical usefulness.
metadata, the decision support of DW is under the control of technical users. As DW evolves, extracting data from OLTP systems to DW becomes more complex. If metadata is integrated with DW, the extraction can be automatic. When metadata is separated from the DW development process, many tools can have a set of mutually exclusive metadata that describes the same data (Anahory & Murray, 1997).

For this integration, this paper proposes taxonomies for dataflow and metaflow. Furthermore, an architecture for DW is developed on the basis of these taxonomies. The architecture consists of seven components: legacy, extracting, operational data store (ODS), DW, data mart (DM), application, and metadata. The emphasis is on the metadata component, which consists of technical metadata and business metadata. A generic metadata schema is built for DW. Because the data warehousing system is built mainly for business users, the business metadata is further explored. The architecture, with its metadata component, is applied to implementing a real-life DW in order to illustrate its practical usefulness.

DATA WAREHOUSE: SYSTEM, TAXONOMY, AND ARCHITECTURE

Data Warehousing System

Users of data warehousing systems are classified into business users and technical users. Business users include executive users, casual users, business analysts, and power users (Poe, 1997). Technical users include system administrators, application developers, operators, technical supporters, and designers. The framework for data warehousing system may be derived from an input-process-database-output system by Orr (1998). The input corresponds to the operational database. The process corresponds to the extracting software. The database includes informational database which consists of DW, ODS (Inmon et al., 1997), and DM (Hackney, 1997; Hoven, 1998). The output utilizes the database, i.e., application software. A metastore, a database containing metadata, controls and manages the other parts.

The term “data warehousing” is used to emphasize the dynamic characteristics of DW (Hackathorn, 1995). The dynamics of DW are expressed by the following five flows.

Taxonomies for Dataflow and Metaflow

Taxonomies for dataflow and metaflow are important because they can describe all of the processes of using data warehousing system (Lee et al., 1997). Hackathorn (1995) originally investigated five flows on which taxonomies are based: inflow, upflow, downflow, outflow, and metaflow. This paper adopts this classification (Table 1).

Inflow feeds data from legacy systems and other external sources into the informational database or users’ application. In this process, data from the operational system may be transformed. Upon entering the informational database, data become highly detailed. Upflow aggregates and summarizes the highly detailed data. Downflow purges or archives data into storage media such as magnetic tape. It helps maintain the vitality of informational database. Outflow implies that users utilize data from the informational database by a simple or an advanced tool with functions ranging from basic management reporting to complex, drill-down, and analytical processing.

Metadata are related with the other four flows. Metaflow keeps metadata up to date. The metaflow can depend on a particular DBMS (Database Management System), and its content can vary depending on the purpose of its usage.

In this paper, we propose taxonomies for dataflow and metaflow. The taxonomy for dataflow has three dimensions: inflow, ownflow, and outflow. Metaflow can be classified according to three criteria: situation, distribution, and usage. Figure 1 depicts the taxonomy for dataflow. The dimension called inflow implies how to access legacy systems. Inflow can be one of three types (Widom, 1995). First, end-

Table 1: Dataflow and Metaflow

<table>
<thead>
<tr>
<th>Flows</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataflow</td>
<td>Inflow</td>
</tr>
<tr>
<td></td>
<td>Upflow</td>
</tr>
<tr>
<td></td>
<td>Downflow</td>
</tr>
<tr>
<td></td>
<td>Outflow</td>
</tr>
<tr>
<td>Metaflow</td>
<td>Processes that move metadata, which come from several flows.</td>
</tr>
</tbody>
</table>

Figure 1: Taxonomy for Dataflow
9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: www.igi-global.com/article/metadata-oriented-architecture-building-datawarehouse/3269?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Library Science, Information Studies, and Education. Recommend this product to your librarian: www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

A Template-Based Analysis of GRL
www.igi-global.com/chapter/template-based-analysis-grl/4390?camid=4v1a

Enforcing Cardinality Constraints in the ER Model with Integrity Methods
Mira Balaban and Peretz Shoval (2002). *Advanced Topics in Database Research, Volume 1* (pp. 1-16).
www.igi-global.com/chapter/enforcing-cardinality-constraints-model-integrity/4319?camid=4v1a

Agreements Treaties and Negotiated Settlements Database
www.igi-global.com/chapter/agreements-treaties-negotiated-settlements-database/7986?camid=4v1a

Metaschemas for ER, ORM and UML Data Models: A Comparison
www.igi-global.com/article/metaschemas-orm-uml-data-models/3277?camid=4v1a